

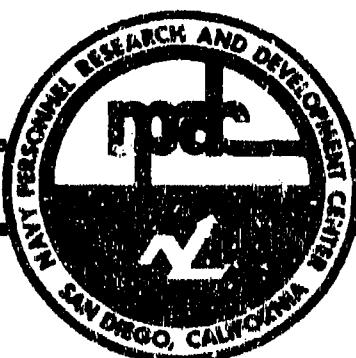
AD A11 3208

NPRDC TR 82-37

MARCH 1982

(12)

PREDICTION OF JOB PERFORMANCE:
REVIEW OF MILITARY STUDIES



NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

DTIC
APR 9 1982
H



82 04 09 023

NPRDC TR 82-37

March 1982
(12)

**PREDICTION OF JOB PERFORMANCE:
REVIEW OF MILITARY STUDIES**

Robert Vineberg
John N. Joyner

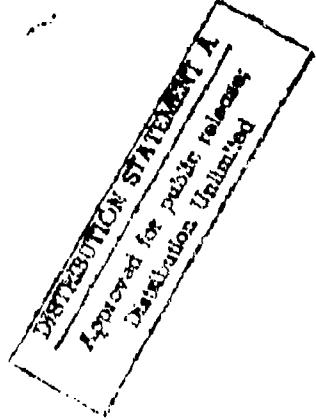
Human Resources Research Organization
Carmel, California 93923

Reviewed by
Martin F. Wiskoff

Released by
James F. Kelly, Jr.
Commanding Officer



Navy Personnel Research and Development Center
San Diego, California 92152



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DD FORM 1 JAN 72 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

the most practical method of objective measurement; (2) Because job sample tests are very expensive to construct and administer, their use is not practical unless the job is extremely costly or critical; (3) Use of supervisors' ratings as the only measure of job performance should be restricted to jobs for which motivation, social skill, and response to situational requirements are the only attributes worth measuring. Two promising approaches to improved prediction are the selective use of miniaturized training and assessment centers and the use of self-paced training performance as a predictor. The review includes abstracts of the studies that were reviewed.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FOREWORD

This effort was conducted under contract MDA 903-80-C0724 with the Human Resources Research Organization in support of work unit ZF63-520-001-030-01.01 (Prediction of Performance). Its purpose was to determine the current state of the art for predicting performance in military jobs. Results are intended for use by fleet personnel concerned with the evaluation of on-the-job performance, by personnel researchers, and by cognizant personnel in the Military Personnel Command.

Appreciation is extended to Dr. Wayne S. Sellman, Office of the Assistant Secretary of Defense (MRA&L), who provided the logistics support for contract initiation. The contracting officer's technical representative was Dr. Charles H. Cory.

JAMES F. KELLY, JR.
Commanding Officer

JAMES J. REGAN
Technical Director

Accession Per	
NTIS GRA&I <input checked="" type="checkbox"/>	
DTIC TAB <input type="checkbox"/>	
Unannounced <input type="checkbox"/>	
Justification _____	
By _____	
Distribution/ _____	
Availability Codes	
Dist	Avail and/or Special
A	



SUMMARY

Problem

The Navy's current personnel assignment system performs well in assigning qualified personnel to technical schools, but it appears to be less adequate in predicting on-the-job performance. In recent years, many efforts have been made to measure and predict military job performance, but these efforts have not been systematically cataloged or reviewed.

Objectives

The objectives of this study were to summarize recent (1952-1980) published research that has measured and predicted performance in military jobs and to provide a systematic report on the current state of the art.

Approach

The literature published between 1952 and 1980 concerned with predicting job performance of enlisted personnel in U.S. military service was reviewed. Performance was selectively defined in terms of job proficiency, job performance, and suitability for military service. Because the definition of performance did not include accomplishments in training as a criterion, much of the information from traditional studies of the validation of selection and classification procedures--the relation between predictor variables and performance in training--was intentionally omitted. The review did not examine prediction of decisions to reenlist or leave the service.

Findings

1. The majority of studies used ratings to assess performance. Tests of job proficiency were used in 18 percent of the studies; job sample tests of performance were used in 13 percent.
2. Aptitude, biographic or demographic variables, and interest or attitude variables were the most frequently used predictors of job performance. There was little evidence of change in the pattern of predictor use over time.
3. Correlations were low (generally .00 to .30) between job sample tests and paper-and-pencil job knowledge tests and between global ratings and job sample tests. Correlations between ratings and job knowledge tests were sometimes slightly higher, but rarely exceeded .35.
4. In most studies, aptitude variables predominated as predictors. Job knowledge tests and job sample tests had median correlations of .40 and .31 with aptitude predictors. Composite measures of suitability had a median correlation of .24, and global ratings of performance were the least predictable, with an average correlation of .15.
5. Where sufficient data were available for separate analysis of specific predictor-criterion combinations, median correlations for aptitude with job knowledge ranged from .30 to .50; for training grade with job knowledge, .40 to .50; for aptitude with job sample tests, .10 to .35; and for aptitude and biographic/demographic predictors with supervisor ratings, .12 to .17. Earlier performance in training correlated with supervisor ratings with a median validity of .23.

Conclusions

1. For a majority of jobs, job knowledge tests provide the most practical method of objective measurement. They are much less expensive to construct and to administer than are job sample tests. They are more predictable and more suitable for jobs in which incumbents are widely dispersed. However, the variability of the validity coefficients reported in the literature for job knowledge tests suggests that proper psychometric procedures must be followed in their construction. There is evidence that considerably higher correlations can be obtained if the knowledge tests are constructed on the basis of careful analysis of job behavior.
2. Because of the high expense of developing job sample tests for validation, their use is impractical except where a job is extremely critical or costly.
3. Supervisors' ratings are of dubious value for several reasons, including (a) the phenomenon of halo that makes them unsuitable for the specificity implied in measuring technical performance and (b) the frequent lack of familiarity on the part of raters regarding the work of the persons they evaluate.

Recommendations

1. Greater attention should be given to characteristics of jobs and people in jobs when conducting test validation studies. For example, the predictability of a criterion depends on many factors besides the prediction-criterion relationships. Some of these factors are variability of performance across job holders, job difficulty, performance levels at entry and after various lengths of time, the effective ceiling in job performance, and how soon and by what percentage of incumbents the ceiling is reached. Greater understanding of the predictability of job performance criteria will require systematic study of these previously neglected factors in conjunction with the predictor-criterion relationships.
2. The use of miniaturized training and assessment centers warrants further evaluation, especially in predicting performance for demanding jobs where the size of the training investment may warrant the added costs of prediction.
3. Relationships among predictor, training, and job performance variables must be better understood. In this context, there should be a focus in self-paced training on variables that can serve as supplemental predictors to entry tests.
4. Use of supervisors' ratings as the sole measure of job performance should be restricted to jobs for which motivation, social skills, and response to situational requirements are the only attributes worth measuring.

CONTENTS

	Page
INTRODUCTION	1
Problem	1
Objectives	1
Background	1
APPROACH	2
FINDINGS	3
Literature Review	3
Criterion Variables	3
Predictor Variables	7
Predictor-Criterion Relationships	10
Prediction of Performance for Specific Subgroups	17
Prediction of Performance by the ASVAB	19
DISCUSSION AND CONCLUSIONS	21
Major Patterns of Use	21
Relationships Among Predictors and Criteria	21
Current Predictive Validities	21
Possible Improvement in Predictive Validities	22
Reliance on Global Ratings	22
Job Sample Tests	23
Job Knowledge Tests	23
Supervisor's Global Ratings	23
Two Promising Approaches for Future Validation Development	23
RECOMMENDATIONS	24
REFERENCES	27
BIBLIOGRAPHY	35
APPENDIX--ABSTRACTS OF PUBLISHED STUDIES	A-0
DISTRIBUTION LIST	

LIST OF TABLES

INTRODUCTION

Problem

The fact that military recruits are selected and classified for particular kinds of military training and jobs implies that their subsequent job behavior can be predicted. However, personnel selection procedures in the past have usually been validated against training criteria rather than against job performance. The Navy's current personnel assignment system performs well in assigning qualified personnel to technical schools, but it appears to be less adequate in predicting on-the-job performance. In recent years, many efforts have been made to measure and predict military job performance, but these efforts have not been systematically cataloged or reviewed.

Objectives

The objectives of this study were to summarize recent (1952-1980) published research that has measured and predicted performance in military jobs and to provide a systematic report on the current state of the art.

Background

In the past, military job selection procedures have usually been based on training criteria, rather than on job performance criteria, for several reasons. Some of these reasons are:

1. Measurement of job proficiency and behavior is difficult. Factors contributing to this difficulty include (a) the differences in requirements across billets, (b) the changes in demands that go with increased job experience, (c) the difficulty of standardizing the experiences of incumbents, (d) the scarcity of measurable products based on job performance, and (e) the absence of objective means for describing many aspects of performance.
2. Measurement of job proficiency or performance may not reveal sufficient differences among incumbents to test selection and classification strategies. After selection and training have filtered out very poor performers, both the range of predictive characteristics and of incumbents' performance become restricted. In addition, this restriction in range is intensified where job demands themselves are not great.
3. Performance in training provides evidence of necessary constituents of job performance. Training makes an obvious contribution to job performance because a person must know what to do and be able to do it before performing a job. Much of the information and skill that affect performance in military jobs is acquired through technical training. Therefore, knowledge and skill demonstrated in training are regarded as evidence of a person's ability to perform on the job.
4. Training performance provides a measure of learning ability, which is usually a prerequisite for future growth and performance. Performance during training provides information about a person's sheer ability and desire to learn the skills that the job will require. The capacity to learn what is demonstrated during training may be a more important contributor to--and predictor of--later job success than the particular type of training received.

In spite of these influences, the military services have increasingly attempted to validate selection procedures against some more ultimate criterion of performance than training for the following reasons:

1. Social and manpower policies have called for closer attention to the true minimum requirements of acceptable job performance so that enlistment standards can be set at appropriate levels.

2. Sensitivity has developed to potential bias in the use of selection tests and to the need to validate them against job behavior or measures of proficiency that accurately reflect job behavior.

3. "Systems analysis" has been applied to manpower issues, leading to an effort to model all elements of the selection-training-performance sequence.

4. Training has been implemented to a fixed (hence, frequently nondifferentiating) criterion, and training methods that reduce identifiable differences in training performance (e.g., self-paced instruction) have been used.

Thus, the Army, for example, in an extensive program, has developed skill qualification tests (SQT) that contain several direct measures of performance (hands-on component, job-site component). These tests are used both to diagnose training needs and to establish soldiers' eligibility for promotion. They have also provided criterion measures that are currently being used in validation studies of the Armed Services Vocational Aptitude Battery (ASVAB). The Navy has validated selection tests against performance in several jobs, including those of recruiter, sonar technician, and Marine Corps drill instructor, and is currently constructing performance tests to be used in ASVAB validation studies in three Marine Corps military occupational specialties (MOSs).

APPROACH

Reports published between 1952 and 1980, concerned with the prediction of job performance of enlisted personnel in the U.S. military establishment, were reviewed and summarized (see the appendix for abstracts).

On the criterion side, the definition of performance was restricted to variables that reflect how well an individual performs in service. These include work-related measures (e.g., technical performance, proficiency) and measures of suitability to military service, recognizing that military performance is generally considered to extend beyond particular job duties. Reenlistment was not included as a criterion variable.

On the prediction side, only research in which the predictor variables have potential use in selecting and classifying personnel were reviewed, thus restricting predictors to characteristics directly associated with individuals (psychological test scores, biographical information, training and work achievement, attitudinal information including perceived characteristics of the work environment, trait ratings). Situational variables, such as organizational and supervisory atmosphere and situational stress, were excluded. The review did not examine prediction of decisions to reenlist or leave the service.

While the review included military training, performance was not treated as a criterion. Thus, much of the information contained in traditional validation studies was intentionally omitted. Nonempirical articles and methodological studies have generally been omitted as well, except for pertinent review articles and symposia proceedings sponsored by the military departments. Such reports often include both descriptions of experimental studies and discussions of current issues.

A literature search identified the following types of published research:

1. Bibliographies and annotated bibliographies for military personnel, training, and human factors research laboratories, 1952-1980 (including those for existing laboratories and their predecessors, as well as former laboratories).
2. Proceedings of annual conferences of the Military Testing Association, 1967-1980.
3. Proceedings of relevant symposia conducted by human research laboratories and other groups (Mullins & Winn, 1979; Pope & Meister, 1977).
4. Annual Review of Psychology, 1954-1979 (articles entitled "Industrial Psychology," "Personnel Selection," "Personnel Management," "Psychology of Men at Work," and "Personnel Selection and Classification Systems").

Information about current work was obtained from the human research laboratory for each service, as well as from organizations such as the Center for Naval Analysis. Additional sources of ongoing research sponsored by the U.S. military services were obtained from a computer search of the Research and Development Information System (RDIS), an automated data base maintained by the Navy Personnel Research and Development Center (NAVPERSRANDCEN).

In summarizing and comparing information about these studies, the review of literature focused particularly on the kinds of criterion and predictor variables used in the research, the levels of predictive accuracy attained, and the major issues in the prediction of job performance. The reports abstracted in the appendix contain predictive validation studies. Additional reports on related topics are cited in the text but not in the appendix because they contain no specific validation studies.

FINDINGS

Literature Review

Criterion Variables

Published literature that reports empirical relationships between predictors and measures of job performance has been categorized by criterion domain and presented in Table 1. The job proficiency domain has been divided into measures of job knowledge, such as the paper-and-pencil tests often used in the services to determine a person's eligibility for promotion; measures of task performance, which simulate complete job tasks; and measures of task element performance, which simulate components of tasks.¹ The job performance domain has been divided into global ratings of performance, job element and task level ratings of performance, measures of productivity, and grade or skill level attained.

¹Measures of task performance and task element performance differ in degree of task completeness or fidelity. A performance test (job sample test) that represents an entire task would be coded in the task performance category even though some of its aspects like initiating cues may differ from those of the actual job. A test representing only part of a task--for example, ability to make the auditory discriminations required to operate sonar equipment, but not the manual manipulations--would be coded in the task element performance category.

Table I
Publications Categorized by Criterion Domain (1952-1980)

Publication Year	Job Proficiency			Job Performance			Grade or Skill Level Attained	Suitability
	Knowledge	Task Performance	Task Element	Global Rating	Job Element	Productivity		
1952				4	46,5	139		
1953	27,139	27,139	27	27,139	119		7	
1954								
1955				9,115,116				
1956				2,137				
1957			3	35,58,59,60,61,66				
1958				71,107,108,134,135				
				70,87,92,114				
1959	15,16,17	87	8	113,142				
1960	67		91	56,69,74				
1961			90	73				
1962								
1963				127,132				
1964				55,75,95				
1965				129				
1966								
1967				128				
1968				29				
1969				50,106				
1970				28,37,120				
1971	28			122				
1972	122			122				
1973				1				
1974				126				
1975	62		112	62,131				
1976				23,24,25				
1977				22,30,102,133				
1978	85	32,85	32	133				
1979		33,34	12,89	121				
1980		10,31	26,117	19				
				12				
							26	26

Note: Numbers refer to report abstracts in the appendix.

The suitability domain has not been subdivided. Most studies of overall adaptation to military service, such as attrition studies, use a composite criterion of two or more of the following measures: completion of term of enlistment, recommended eligibility for reenlistment, incidence of misconduct, advancement in grade or skill level, performance ratings, and type of discharge.

Overall, the majority of studies used global ratings of performance as a criterion; second in frequency of use were measures of suitability. Of 114 published studies, 48 percent used ratings alone as a criterion; 30 percent used a measure of suitability exclusively. In contrast, only 21 studies (18%) reported using a measure of job proficiency and in only 15 of these (13%) did the test of proficiency involve actual performance.

In the proficiency and performance domains, there have been no major shifts in criterion use. In about the last decade, however, there has been some proportional increase in the use of performance testing: in the 1950s and 1960s, the ratio of studies using ratings to those using knowledge tests and those using performance tests was about 12 to 2 to 1; between 1970 and 1980, the ratio was about 6 to 1 to 2.

Job proficiency, as mentioned previously, refers to the skill and knowledge needed for job performance, while job performance refers to actual job behavior. The contrast between these two domains has been referred to as the proficiency/performance distinction--the distinction between what a person knows or can do (usually as demonstrated on a test) and what a person does (usually as observed on the job). Typically, measures of proficiency focus on how a job should be performed, while measures of job performance focus on how a job is performed. The distinction between these domains and measures, as well as their characteristics, have been described elsewhere (e.g., Alluisi, 1977; Schultz & Siegel, 1961; Thorndike, 1949) and are generally known.

Because proficiency is usually assessed by an achievement test, its measurement can possess considerable objectivity and reliability. However, proficiency measures are limited in scope. They do not provide direct information about an incumbent's motivation for performing or about actual performance.

Job performance, on the other hand, is usually measured through some form of rating. Ratings offer the possibility of taking account of the multidimensional² character of performance. Although there are alternatives to ratings, their applications are limited. Measurements of system performance, for example, are generally not a reliable basis from which to make inferences about individual performance. Product and output measurement is restricted to jobs that have products to evaluate or output to count. Further, objective evaluation of products requires a consensus about the appropriateness of the objective measures.

While ratings are understandably the most frequently used measure for assessing performance, their vulnerability to bias and the effects of halo are well known (Landy & Farr, 1980; Thorndike, 1949; Wherry, 1950). Guion (1978) has noted that job requirements

²There is fairly universal agreement today that job performance is complex and that its dimensions may be related to a greater or lesser degree. For a classic discussion of the issues of dimensionality at a moment in time, over time, and across individuals, see Ghiselli (1956a, 1956b). For discussion of the issues involved in combining dimensions into a single construct or measuring them separately, see Seashore, Indik, and Georgopoulos (1960) and Schmidt and Kaplan (1971).

may be "partially defined by styles of the people who hold them." It is clear that supervisors similarly define the performance requirements for the persons they rate. Dunnette and Borman (1979) have indicated three classes of factors that influence performance ratings: rater's organizational level, rater's characteristics, and ratee characteristics.

The ability of raters to differentiate among aspects of job performance (discriminant validity) is open to question. Hausman and Strupp (1955) provided evidence that ratings of technical performance are contaminated by nontechnical factors. Vineberg and Joyner (1978) found that raters showed "little discrimination among different aspects of job performance" when worker-oriented task level ratings were used to evaluate performance in a variety of Navy jobs. Dunnette and Borman (1979) cited several studies suggesting the "inability of raters to go beyond a certain level of precision in making their observations and recording their ratings."

Another crucial, but generally neglected, aspect of ratings is the degree of familiarity of raters with the work of persons they evaluate. For two Air Force jobs, Wiley (1975a) found that "other" supervisors did not differ from immediate supervisors in their degree of agreement with subordinates regarding tasks performed. Jobs offer different opportunities for supervisors to observe ratees, and Wiley and Hahn (1977) provided evidence that these differences affect the number of significant correlations that are obtained between predictors and ratings of performance. Wilson, Mackie, and Buckner (1954) found that the number of ratings made depended on whether the rater was an officer or petty officer. They also reported that most raters in a shipboard situation did not have an opportunity to observe ratees in the performance of enough tasks.

As mentioned earlier, different measures can be used to assess job proficiency and job performance. Although the relationship among measures is obviously of interest, unfortunately, such information is limited. Ratings are frequently used in job settings and, to a lesser degree, in school settings, but more objective measures to which they might be compared are not. Performance tests are rarely used. When they are used, they are most often administered in schools to persons in the latter phases of training or to recent graduates.

The extent to which performance test scores of trainees and relatively inexperienced job incumbents can be used to estimate predictive relationships for experienced job incumbents is an open question. There is evidence, for example, that performance-mediating factors change as learning occurs (Fleishman & Fruchter, 1960; Fleishman & Hempel, 1956). Accordingly, factors that predict performance after time on the job could be quite different from those that predict near-term or initial performance.

In any event, correlations between job sample tests of proficiency and paper-and-pencil tests of job knowledge have generally been low, ranging from .00 to about .30 (Crowder, Morrison, & Demaree, 1954; Engel & Rehder, 1970; Mackie, Ridhalgh, & Schultz, 1978; Shirkey, 1965, 1966; Urry, Shirkey, & Waldkötter, 1965; Yellen, 1966).³

³Knowledge tests and job sample tests usually measure different components of performance, and knowledge tests often measure "very little of whatever a person gains by time on the job" (Tidy, 1960). One reason is that knowledge tests are frequently developed by school personnel and their content is often derived from training materials. Such content may focus not on concrete elements of performance (cues, behavior), but on general descriptions of procedures or on theoretical, terminological, and historical information not directly descriptive of performance. If knowledge test content were to be derived through job analysis procedures that focus on behavior--such as Flanagan's (1954) critical incident method--they might correlate more highly with job sample tests, discriminate more readily among incumbents who are perceived as effective and ineffective, and share more variance with job experience variables.

Occasionally, however, higher correlations have been obtained. Mackie, Wilson, and Buckner (1954) obtained correlations of .46 and .35 between job sample and knowledge tests in two Navy jobs. Grings and Rigney (1953) obtained a correlation of .69 between a knowledge test and a test of trouble-shooting performance. In four Army jobs, Vineberg and Taylor (1972b) constructed knowledge tests that assessed only information directly relevant to job performance and obtained correlations of .58, .59, .68, and .78 with lengthy performance tests.

As for ratings, their low reliability limits their correlation with any other measure, regardless of the amount of variance they may share. Correlations between ratings and job sample tests of proficiency have been low, with only an occasional correlation appearing above .30 (Crowder et al., 1954; Engel & Rehder, 1970; Mackie & High, 1959; Mackie et al., 1978; Mackie, Wilson, & Buckner, 1954; Vineberg & Taylor, 1972b). Findings in the combined industrial and military literature are similar. Severin (1952) reported median correlations of .23 for proficiency tests and supervisor ratings, and of .32 for proficiency tests and associate ratings.

Correlations between ratings and knowledge tests were perhaps slightly higher than those between ratings and performance tests, but they rarely exceeded .35 (Crowder et al., 1954; Engel & Rehder, 1970; Mackie et al., 1954; Vineberg & Taylor, 1972b). Merenda (1959) summarized relationships between performance ratings and job knowledge tests (promotion examinations) in 40 petty officer rates. Median correlations for petty officers 2nd class (13 jobs) and 3rd class (18 jobs) were .21 and .25 respectively. For petty officers 1st class (9 jobs), a considerably higher value of .49 was obtained.

The low relationship among measures in the proficiency and performance domains clearly indicates that they cannot be substituted for each other. Indeed, to substitute one measure for another would require more than high intercorrelation (Smith, 1976).

Predictor Variables

Table 2 provides a very rough picture of the power of the predictive relationships that have been obtained with criterion measures in selected domains. Distributions of validity coefficients^{*} were assembled for job knowledge tests and global ratings of performance (for which data are abundant) and for task performance tests and suitability criteria (for which data are only marginally sufficient). Insufficient data were available to construct distributions for the remaining measures. It must be remembered that the relationships are determined in part by the reliability of the criteria with which predictors have been coupled. Those associated with more reliable criteria will, on the average, demonstrate higher predictive potential. Also, Table 2 is based on a mix of validities (e.g., Armed Forces Qualification Test (AFQT), aptitude index, and selector scores). Distributions of validities for single predictors would have been somewhat lower.

In extracting the coefficients from the published literature, a variety of additional problems were encountered: Some studies give zero-order correlations and multiple correlations, while others omit the zero-order correlations. Some studies correct validities for the effects of selection; others do not. Some report cross-validities; others do not. To summarize the coefficients inevitably required some degree of arbitrariness, compromise, and selectivity.

*For a discussion of the limitations of validity coefficients in describing the usefulness of predictor-criterion relationships, see Taylor and Russell (1939) and Brogden (1946).

Table 2
Distribution of Validity Coefficients for
Four Criterion Measures

Range of Validity Coefficients		Job Knowledge	Task Performance	Global Rating	Suitability	Total
From	To					
.65	.69		1			1
.60	.64	2	1			3
.55	.59	10		1		11
.50	.54	18		2		20
.45	.49	10		3		13
.40	.44	19	2	1		22
.35	.39	14	3	6		23
.30	.34	13	4	17	3	37
.25	.29	13		17	6	36
.20	.24	1	1	26	3	31
.15	.19	5	3	33	4	45
.10	.14	1	2	35	2	40
.05	.09	2		26	1	29
.00	.04	2	1	23		26
-.05	-.01			10		10
-.10	-.06			2		2
-.15	-.11			2		2
Number		110	18	204	19	351
Median Coefficient		.40	.31	.15	.24	

In general, the following principles were followed in assembling the distributions in Table 2 and in later tables:

1. To improve the analysis, only zero-order correlations were used.
2. Only validities for operational predictors or cross-validities for experimental predictors were used.
3. Coefficients for experimental predictors that had not been cross-validated were omitted.
4. Median values were used when separate validities were reported for several samples in the same job, as well as for validities reported for several methods of predictor item selection or weighting.
5. Correlations for separate subgroups (e.g., mental categories I-III vs. category IV, black vs. white) were not included. Sample sizes in almost all studies were large enough that correlations of .30 were significant beyond the .05 level.

The distributions in Table 2 suggest that, overall, job knowledge tests of proficiency are predicted best, global ratings of performance are predicted worst, and task performance tests and suitability criteria occupy intermediate positions. The validities for proficiency criteria (tests of knowledge and task performance) and suitability criteria are high enough to be of considerable practical value in the selection of military personnel. The apparently greater predictability of knowledge tests is probably attributable partly to a relatively large representation of composite predictors for the validities that have been reported for this measure, partly to the homogeneity and reliability of knowledge tests, and partly to the verbal and cognitive requirements that they share with aptitude predictors.

The low validity of global ratings of performance is consistent with Ghiselli's (1966) findings that the general validity of aptitude tests for predicting performance in industrial jobs was about .20 where performance criteria were likely to be ratings, even though productivity, errors, and accidents were also included. Ghiselli contrasted this level of validity with a coefficient of the general order of .30 for training criteria. Presumably (although Ghiselli did not explicitly say so), such criteria included knowledge tests, for which an average validity of .40 has been obtained.

Predictor variables used in the previously identified studies were categorized by publication year and predictor type. Table 3 presents the frequency of predictor variables used in 109 studies reported between 1952 and 1980. There was no evidence of changing trends in the pattern of predictor use, except that the three studies predicting performance on the basis of assessment center evaluations or miniaturized training all occurred in the 1970s (Cory, in press; Dyer & Hilligoss, 1977; Siegel & Bergman, 1972; Siegel, Bergman, & Lambert, 1973; Siegel & Leahy, 1974; Siegel & Wiesen, 1977).

Table 3
Frequency of Predictor Variables Used in 109 Studies
(1952-1980)

Measure	Frequency
Aptitude	69
Biographic/demographic	48
Interest, attitude, self-description	28
Personality test	13
Assessment center evaluation and miniaturized training performance	3
Training	18
Instructor or peer rating	8
Test scores and grades	12
Performance simulation	4
Job element and trait ratings	10
Experimental	23
Physical and physiological	4
Job satisfaction	2
Preservice arrests and service-incurred disciplinary actions	5
Grade or job type level	3

In most studies, aptitude variables and at least one other type of predictor were used. Biographic/demographic predictors, when used, included age and years of education. Experimental measures include those believed to assess a variety of different constructs--cognitive, perceptual, risk-taking behavior, persistence--for which the definitions vary greatly in clarity.

Predictor-Criterion Relationships

Sufficient information to describe particular predictor-criterion relationships is available only for some classes of variables. The following discussion is organized in terms of criterion domains and the various criteria within each domain. No attempt has been made to discuss relationships among all possible combinations of predictors and criteria.

Job Knowledge Criteria. Knowledge tests are typically used to assess achievement and to diagnose student deficiencies during training. Less often, they are administered to job incumbents in determining eligibility for promotion. Although, in this instance, they may serve as an alternate criterion to course grades for validating classification tests, they are not, of course, a direct criterion of job performance.

Where information has been reported about the relation of aptitude scores to knowledge test scores for job incumbents, the size of the relationship depends on several factors, including characteristics of the aptitude measure, the job, and the knowledge test. Composites of aptitude scores (e.g., aptitude indices, AFQT) have been found to have maximum corrected validities as high as the mid .70s. Validities for single tests tend to fall considerably lower. For example, Brokaw (1959a, 1959b, 1959c) obtained validities in 46 Air Force specialties ranging from .19 to .75 (median .58) for aptitude indices of the airman classification battery with the airman proficiency test. He obtained validities for AFQT with the criterion ranging from .07 to .56 (median .38). Similarly, Vineberg and Taylor (1972a) reported a median correlation of .41 between AFQT and job knowledge in four Army jobs. Curtis (1971) obtained average validities of about .31 (ranging from -.03 to .66) for selectors for 19 Navy training courses against later performance of graduates on advancement examinations. Individual tests of the Navy basic test battery and the factor reference battery (Morsh, 1957) had average validities of about .22 (ranging from -.24 to .73) and .14 (ranging from -.13 to .57), respectively, against the same criterion.

Treatment of relationships among particular aptitudes and performance on knowledge tests of incumbents in particular jobs is beyond the scope of this review. Such aptitudes as those tapped by verbal ability tests of word knowledge, mechanical knowledge, and arithmetic have consistently demonstrated sizable validities across jobs, but they have usually been based on analyses that use school grade as a criterion (Gragg & Gordon, 1951; Tupes, Brokaw, & Kaplan, 1960). Where the knowledge tests have been administered to job incumbents rather than trainees, however, similar validities have generally been obtained (Brokaw, 1960).⁵

⁵Brokaw obtained a median validity of .57 for aptitude indices and final school grade suggesting that aptitude has about the same validity for the prediction of either school grade or job knowledge. Ghiselli (1966) found correlations of aptitude and training to be about 10 points higher than correlations of aptitude and job performance.

Although biographical information has generally been viewed as one of the best predictors of performance (Dunnette & Borman, 1979; Ghiselli, 1966; Taylor & Nevis, 1961), it has typically been used to predict criteria other than job knowledge tests. For example, it has frequently been used in conjunction with other variables to predict attrition (e.g., Erwin & Herring, 1977; Guinn, Kantor, Magness, & Leisey, 1977; Holberg & Pugh, 1977). What information is available, however, indicates both the validity of biographical material against a measure of job knowledge and the significant contribution it makes when used in combination with aptitude predictors. Brokaw (1960) reported biographical information among the best of four predictors in the airman classification battery in about half the criterion groups for mechanical, administrative, and electronics specialties.

The role of a particular biographic/demographic variable, length of formal education, is worthy of mention. Judy (1960) cited several studies in which formal education generally was found to have a low relationship to knowledge test performance--zero-order correlations ranged from -.02 to .25, with a median of .17. Similar findings were reported by Vineberg and Taylor (1972a), who obtained correlations of .01, .12, .14, and .15 in four jobs. Higher correlations, Judy indicated, would suggest that irrelevant academic factors were playing an important role in the measurement of performance. (In contrast to its poor predictive validity for job knowledge, length of education has been found to make a significant contribution to the prediction of other criteria, most notably composite measures of suitability.)

Job experience--months on the job, months in grade, the number of tasks, and the difficulty of those tasks--has been studied as another variable. Judy (1960) reported that the proportion of variance attributable to these experience variables in performance on a job knowledge test for mechanics was significant but quite low. He concluded that job knowledge tests "in the mechanical area measure very little of whatever a person gains by time on the job . . . it is doubtful that a general examination (covering a whole specialty) should be expected to discriminate high- and low-experience groups" (p. 4).

In discussing low predictive relationships with performance measures taken in the job situation, Christal (1979, pp. 131-145) found little reason for higher validities. If the selection process picks the correct people to train and the training is efficient, job experience should not contribute greatly to differences in technical skill and predictors (aptitude) will not necessarily be highly related to job performance.

It remains an open question whether findings such as Judy's (1960) have been obtained because knowledge tests fail to measure adequately what has been learned, because incumbents are all very similar after training, or both. It may be noted that Vineberg and Taylor (1972a) found months on the job (job learning) to be the most important predictor of performance on job knowledge tests (correlations of .55, .45, .63, and .46 in four jobs). Adequate control of the amount of learning that has occurred, both before job entry and on the job, is a major problem that will be discussed later.

Studies of relationships among attitudinal and interest variables and objective measures of proficiency, such as job knowledge tests, are rare in the military literature. Hickerson, Hazel, and Ward (1975) cited Tuttle and Hazel (1974), who suggested that the military's primary concern with job satisfaction is related to motivation and career intent rather than to performance. Ratings of job interest and perceived utilization of talent and training were found to be unrelated to performance on job knowledge tests in seven Air Force specialties. Hickerson et al. have provided analysis of methodologies for establishing relationships between performance and job satisfaction, as well as a review and bibliography of this specialized topic.

Information about the relation of school performance to subsequent performance on job knowledge tests is available from a limited number of studies. Brokaw (1959b) obtained a median correlation of .54 (range from .24 to .80) between final school grade and knowledge test score (airman proficiency test) for 46 specialties. Austin (1955) obtained a median correlation of .52 for 15 specialties, and Curtis (1971) obtained average validities of about .46 (range from .13 to .67) for incumbents in 19 specialties.⁶ The relationship between final grade and job knowledge of incumbents is, on the average, perhaps slightly higher than that between aptitude and job knowledge, which also tends to have a wider range of validities. The validities for the two predictors, however, are similar enough to suggest that performance on paper-and-pencil tests is a common mediating element.

Task Performance Criteria. The cost of administering task performance tests to job incumbents (as distinct from administering them to persons in a school environment) has drastically limited the number of studies in which predictor-performance relationships have been examined. Crowder et al. (1954), in an early study of radar mechanics, obtained a median correlation of .07 between aptitude scores and trouble-shooting performance. Mackie and High (1959) administered performance tests to Navy machinery repairmen and found a median validity of .21 for aptitude scores and a multiple correlation of .39. Earlier school performance on work sample tests correlated .18 with the criterion and had a multiple correlation of .44. Instructors' ratings of school performance correlated .32 with the criterion.

Vineberg and Taylor (1972a) obtained zero-order validities for AFQT with performance on job sample tests of .27, .30, .35, and .35 among incumbents whose job experience ranged over 20 years. Correlations of the criterion with experience (months on the job) were .69, .43, .43, and .39.

In studies of selected aptitude tests and a performance battery, which included job knowledge tests, Mackie and his associates (Mackie, Wilson, & Buckner, 1954; Wilson & Mackie, 1952) reported median zero-order validities of .35 and .17 and multiple correlations of .62 and .56. The earlier school standing of job incumbents (based on achievement test score and instructor ratings) produced multiple correlations of .40 with the criterion in both jobs.

Mackie et al. (1978) administered task performance tests (target detection, report timeliness, target tracking, etc.) to students during the last stages of sonar operator training. Validities for ASVAB tests were quite low, ranging from -.24 to .13 with a median of about .00. A biographical inventory scale correlated -.13 with the criterion. Two measures of school performance--written test average and practical factors ratings --correlated .13 and .34 with the criterion. The study included a variety of other predictors, including other aptitude tests, personality tests, experimental tests of perceptual skills, and physiological measures.

Eaton (1978) used aptitude tests, performance on a training simulator, and job sample test components as predictors of gunnery task performance. Although some substantial validities were obtained (zero-order validities as high as .49 and a shrunken multiple correlation of .56), the sample was small--less than 40 crewmen in a position. When the same predictor and criterion variables were used in a larger sample (Eaton, Bessemer, & Kristiansen, 1979), none of the relationships was confirmed.

⁶Brokaw indicated that validities had been corrected for restriction of range, but Curtis did not; this difference may account for the difference in size of their average correlations.

In another recent study in which aptitude and performance on elements of training simulators were used to predict gunnery performance of tank crewmen (Eaton, Johnson, & Black, 1980 draft), aptitude-derived measures (ASVAB) were not found to be related to performance. Some promising relationships, however, were found between elements of simulated performance and gunnery performance.

Black (1980) reported validity coefficients of .44, .40 and .17 for the combat aptitude area composite (CO) of the ASVAB with performance tests for tank crew gunners, loaders, and drivers. The predictor had shown inversions of these relationships with the criterion (correlations of .13, .23, and .36) when administered to the same job incumbents 4 to 8 months earlier. "Shifts" of this sort in a single study, however, must be viewed with caution. Because such findings can occur partially as a consequence of ceiling effects, it would be desirable to have additional information about the range of performance of different aptitude groups at the end of training and on the job before interpreting these data. For example, do the correlations increase for gunners and loaders because their tasks are relatively more difficult than those of drivers, because their proficiency is low at the end of training, or because most skill is acquired on the job? Does the correlation decrease for drivers because most learning occurs during training or because few differences remain among incumbents after a brief time in the job?

In summary, when task performance tests are the criterion measure, aptitude tests appear to have zero-order validities largely in the .20 to .40 range for the prediction of proficiency. Aptitude composites tend to fall at the upper end of that range, and single tests tend to fall at the bottom. When proficiency is measured by tests of job knowledge, aptitude tests tend to have a somewhat broader range of validities. Zero-order validities range from about .15 to .70, again depending on whether the aptitude score is a composite or a single test. On the average, validities for knowledge tests are about 15 to 20 points higher than for performance tests.

Global Rating Criteria. Table 4 presents predictive relationships that have been obtained for different types of predictors with global ratings of job performance. Few correlations were reported for biographic/demographic, education, interest/attitude, training, and trait rating variables as predictors, and the trends shown are likely to fluctuate with the addition of more cases. Correlations for level of education are shown separately for cases in which it was reported apart from other biographic/demographic variables.

Aptitude, education, and interest/attitude measures all appear to be correlated with global ratings at about the same low level. There is some question, however, about the contribution to be expected from education and interest/attitude variables. Although level of education has been found in the past⁷ to be the best single predictor of suitability and attrition criteria, level of education, particularly high school graduation, has occasionally been found to operate as a suppressor variable. For example, Brokaw (1960) reported negative beta weights and positive validities for education in the prediction of knowledge test scores and suggested that education may make a negative contribution to multiple correlation with aptitude and biographical measures. Harding and Bottenberg (1961) found that such attitude variables as satisfaction with the Air Force, supervisor,

⁷Changes in secondary education policy in recent years may have reduced the validity of high school graduation as a predictor of suitability for the military services.

and job make insignificant contributions to correlations of status variables (e.g., rank, length of service, kind of work performed) with supervisor ratings. They suggest that much of the variance that has been attributed to attitude variables is held in common with more easily specified status variables.

Table 4
Distribution of Validity Coefficients for
Predictors of Global Ratings of Performance

Range of Validity Coefficients		Aptitude	Bio- graphic/ Demo- graphic	Educa- tion	Interest/ Attitude	Training Perfor- mance	Totals for Global Ratings	Concurrent Trait Ratings ^a
From	To							
.80	.84						3	
.75	.79						1	
.70	.74						3	
.65	.69							
.60	.64							
.55	.59					1	1	2
.50	.54	1	1				2	1
.45	.49					3	3	3
.40	.44	1					1	
.35	.39	2		1		3	6	
.30	.34	5	1	2		9	17	
.25	.29	6		2	1	8	17	
.20	.24	12	2	2	1	8	25	
.15	.19	15	4	4	3	9	35	
.10	.14	20	2	2	3	7	34	
.05	.09	15	1	7		8	31	
.00	.04	14		4	5	3	26	
-.05	-.01	8	1	1			10	
-.10	-.06	1			1		2	
-.15	-.11	1			1		2	
Number		101	12	25	15	59	212	13
Median Coefficient		.12	.17	.12	.12	.23		.71

^aConcurrent trait ratings have been presented for comparison with global ratings.

Biographic/demographic information has generally been viewed as one of the better predictors of job performance. In his review of validity studies, Ghiselli (1966) suggested that it was the most successful predictor for both training and job proficiency criteria. Asher (1972), and Asher and Sciarri (1974) found that biographic information has the highest predictive validity when job proficiency is the criterion. In the present review, limited to military studies, there is evidence that biographic/demographic variables may predict only slightly better than aptitude variables.

Of all the predictors reported, performance in training appears to have the highest truly predictive validity (omitting from consideration trait ratings made concurrently with global performance ratings). Validity coefficients for training performance as a predictor are based almost entirely on correlations between final school grade and supervisor ratings made subsequently on the job. Instructor evaluations in training contribute to the computation of school grades to an unknown degree. There is evidence, however, that these instructor evaluations account for the additional potency of performance in training as a predictor. In Table 4, 19 of the 59 validity coefficients for training performance were taken from a study by Curtis (1971). He obtained an average correlation of .21 between final school grade and performance ratings, whereas written test and performance test scores in school had average validities of .16 and .11 with the criterion. Because final grade is often a composite of test scores and instructor evaluations and because test scores alone do not account for the correlations between final school grade and on-the-job ratings, it is reasonable to assume that the addition of instructor evaluations to the composite added about 5 to 10 points to the correlations between school achievement test scores and job performance ratings. Without the hypothesized augmentation by instructor ratings, achievement test scores taken in training predict subsequent ratings of job performance only at about the same level of efficiency as do aptitude, biographic/demographic, and attitude/interest variables.

As in the case of instructor ratings that contribute to prediction of job performance, higher validities may be possible when common or similar elements are present in both the predictor and criterion. Asher (1972) and Asher and Sciarino (1974) have used this so-called point-to-point relationship between such common or similar elements to explain the efficacy of historical information as a predictor of later performance and of "motor" work sample test scores as predictors of job proficiency.

The correlations for concurrent trait ratings reported in the last column of Table 4 are taken from a series of studies by Wiley (1964, 1966, 1974, 1975a, 1976), Wiley and Cagwin (1968), and Wiley and Hahn (1977). They are included in the table to indicate the general level of "predictability" of ratings of overall performance on the basis of trait ratings. The studies used trait ratings primarily as a methodology both to determine job requirements and to analyze rating process requirements. In some of the studies, trait ratings were used to predict performance over time and across installations. The coefficients in the table are cross-validities obtained through cross applications: trait ratings by one supervisor are correlated with overall ratings of performance by another supervisor.

A major contribution of these studies is their focus on the subtleties of the rating process and the information loss that results if ratings are gathered with as little consideration to the characteristics of situations, raters, and ratees, as they often are. Their findings include the following: (1) Patterns and usefulness of trait ratings as predictors vary by skill level, (2) predictability of global ratings varies by skill level, (3) difficult tasks are rated more reliably than easy ones, and (4) ratees must be separated by grade or skill level because ratings rise uniformly by grade and are correlated with skill level.

Suitability Criteria. Prediction of an individual's likely suitability for military service has been studied extensively in the last 30 years (see Table 1). Trends in types of predictors have been described and a listing of publications provided in a review of the related topic--attrition (Wiskoff, Atwater, Houle, & Sinaiko, 1980).

In the studies reviewed, suitability has usually been defined by a criterion that includes measures of a person's availability and continuation of service, the person's

performance during service, or both. These measures include completion of first term of enlistment, eligibility for reenlistment at end of term of enlistment, incidence of misconduct (e.g., court martial, nonjudicial punishment, delinquency and misdemeanors, conviction by civil court, recidivism), advancement in grade or skill level, performance ratings, and type of discharge. Predictors of adjustment have included aptitude and biographic/demographic variables; scores and other information from personality, attitudinal, and interest inventories; psychiatric evaluations; and peer and instructor ratings obtained during early military training.

Three variables--education level, mental ability (aptitude), and age--have consistently demonstrated predictive validity for suitability criteria (Fisher, Ward, & Holdrege, 1960; Flyer, 1960a, 1963, 1964; Gordon & Bottenberg, 1962; Klieger, Dubuisson, & de Jung, 1961; Plag, 1962; Plag & Hardacre, 1964). Table 5 presents a general picture of predictive relationships that have been obtained for suitability criteria. The validity coefficients have been taken from studies in which zero-order correlations for particular predictors have been given. In many of the studies, relationships are reported in expectancy tables rather than as correlations; therefore, they are not readily summarized.

Table 5
Distribution of Validity Coefficients for Predictors
of Behavioral and Performance Suitability Criteria

Range of Validity Coefficients		Aptitude	Age	Education	Ratings	Autobiographical Inventories	Total
From	To						
.45	.49			2			2
.40	.44			1	1		2
.35	.39			3	1		4
.30	.34	2		3	1	2	8
.25	.29	3	4	1	2	2	12
.20	.24	4	1		2		7
.15	.19	1	5				6
.10	.14						
.05	.09	1					1
Number		11	10	10	7	4	42
Median Coefficient		.24	.21	.36	.29	.29	

Multiple validities for education, mental ability, and age range from about .24 to .39 (Lockman, 1974). When it is possible to take advantage of information obtained early in military service--that is, for classification rather than screening--higher validities can be obtained. Flyer (1963), for example, has obtained point biserial validities in the .40s and biserial validities in the .50s and low .60s for predictor composites that include peer and instructor ratings in basic training.

Prediction of Performance for Specific Subgroups

The military prediction literature contains a number of studies in which the performance of selected subgroups is examined. In most of these studies, a demographic variable defines a population of interest, which is often a minority (e.g., mental category IV, blacks, women). Depending on the purpose of the study, comparisons of performance or prediction of performance may be made between the subgroup and some other population. Two special cases are studies concerned with (1) test fairness and (2) the differential validity of selection tests for different subgroups.⁸

Another class of subgroup studies seeks to improve predictability by varying the combinations of predictors (and individuals) according to the characteristics of individuals and jobs. The term "moderator" has sometimes been used to refer to the variables in such studies.⁹

Validities were obtained for a variety of operational and experimental predictors with job performance ratings and progression from apprentice into technical Navy jobs (Cory, 1976a, 1976b; Cory, Neffson, & Rimland, 1980). Subgroups were persons in mental category I-III and IV, as well as blacks and nonblacks. The validities were low because of restriction in range, although most were similar to the average validities reported in Table 4. For example, mental category IV had a median zero-order validity for aptitude scores with rating criteria of about .13 and for educational level with ratings of about .19. Experimental predictors did not add sufficiently to predictive accuracy to warrant operational application. Of interest in Cory's data are the validities that emerged with the more objective criterion of progression to a technical job. Despite the use of a dichotomous criterion, zero-order validities for aptitude scores were in the high .30s and low .40s for nonblack and black groups respectively, and validities for educational level were about .25 for both groups. The Strong Vocational Interest Inventory (SVII) had the highest validities of any variable with the progression criterion.

Vineberg and Taylor (1972a), in the study of four Army jobs discussed earlier, obtained a substantial number of significant and consistent validities for aptitude variables with job knowledge scores for persons in both mental category I-III and mental category IV. No consistency across jobs was obtained for aptitude validities with a job sample criterion.

⁸Differential validity refers to differences in the size of the validity coefficient for different subgroups. Test fairness refers to the absence of selection error for a particular subgroup or its avoidance by means of procedures to adjust characteristics of the prediction equation that would otherwise lead to error. For a description of different models of test fairness see Dunnette and Borman (1979).

⁹Saunders (1956) broadly defined the term "moderator" to refer to "situations in which the predictive validity of some psychological measure varies systematically in accord with some other independent psychological variable." Guion (1967, 1976) described a variety of uses of the term since then: variables used for population definition and control; variables that correlate with correlations (Saunders); variables that correlate with errors of predictions (Ghiselli, 1956); and variables that interact with predictors. He deplored the tendency to mix methodological and mathematical meanings.

Studies concerned with the prediction of military suitability often report validities for selected subgroups (e.g., low-aptitude personnel and high school nongraduates) and occasionally explore the use of different combinations of predictors with these groups. Composites of aptitude, education, and other biographical predictors, for example, have validities of about .30 to .40 for low-aptitude groups with a suitability criterion and zero-order correlations in the .10 to .20 range (Gordon & Flyer, 1962; Plag, Goffman, & Phelan, 1967, 1970).

Some predictors that contribute to prediction of the performance for high school nongraduates are less effective in predicting the performance of graduates. Flyer (1960b) and Gordon and Bottenberg (1962) found that age contributes to the prediction of unsuitability discharges among persons who had not graduated from high school, but they found age unrelated to suitability among graduates. Flyer (1963) reported point biserial validities of .25 and .42 for composites of peer and supervisor ratings in training with a subsequent suitability criterion for high school graduates but higher validities of .41 and .52 for nongraduates. Presumably nongraduates are more heterogeneous.

Kipnis (1961, 1962, 1965) has reported on the interactions among aptitude, experimental trait measures (persistence and insolence), and ratings of performance. Although he sometimes found the evidence inconclusive, intelligence was reported to moderate the validities of the trait measures with relationships between insolence and performance for persons with high aptitudes and between persistence and performance for persons with low aptitudes.

Siegel and his associates (1972, 1973, 1974) used operational Navy instruments and miniaturized job learning situations and tests to predict job sample and rating measures of criterion performance in low-aptitude black and white recruits assigned to machinist mate jobs in the Fleet. Criterion data were obtained 9 and 18 months after job entry. Overall, the operational Navy tests and the miniaturized situations predicted equally well, although the latter tended to be more effective at 18 months. Miniaturized situations seemed to hold promise as predictors, given their potential for additional validity with the use of orthodox test development procedures, as well as their greater face validity, examinee acceptance, and fairness. An interesting aspect of the study was that, after 18 months, there were no significant differences in performance between the low-aptitude sample and a higher aptitude, "A" school-trained, control sample.

Siegel and Wiesen (1977) used a combined assessment center and job learning methodology for the classification of Navy general detail personnel (persons who had not qualified for assignment to a school). Unfortunately, predictions about abilities derived from the experimental classification procedures could not be validated, because later it was found that fleet commands had not followed assignment recommendations (Cory, in press). Ratings of on-the-job performance, job progression (striker/nonstriker status), and retention status were determined 11 and 19 months after assignment. Ratings did not reveal differences between performance of persons who had been recommended for technical jobs and those who had not. In the 11-month follow-up, Cory found significant validity coefficients for classification tests, biographic variables, and assessment center variables in 25, 38, and 21 percent of the cases, respectively. In the 19-month follow-up, these predictors produced 0, 38, and 16 percent significant coefficients. Performance over time was generally predicted best by biographic variables and, next best, by assessment center variables. There was little evidence that the assessment center variables would be of practical use in identifying persons who would progress in grade or remain in the Navy.

Prediction of Performance by the ASVAB

In January 1976, each of the services began using the ASVAB as the only enlisted accessions test both for selection into the service and for job assignment. An Armed Forces Qualification Test (AFQT) score derived from the ASVAB is used to determine enlisted eligibility; aptitude composite scores are derived and used for job placement.

In October 1980, several new forms of the ASVAB (8, 9, and 10) were implemented. Among other differences, the new forms differ from earlier classification instruments in that they omit items that provide information about vocational and other interests of prospective enlistees.

Mackie, Ridhalgh, Seltzer, and Shultz (1980) administered sonar operator job sample tests to a mixed sample of trainees (75% in final stages of training, 25% from other assignments aboard ship) at an antisubmarine warfare training center. The test used high fidelity recordings of sonar signals in a simulation that required "operator responses that are essentially the same as those required . . . aboard ship." Validities for the ASVAB school predictor composite with five subtests ranged from .01 to .60. A multiple correlation between the predictor composite and a performance composite was .43, corrected for restriction in range.

Several studies in which ASVAB composites were used to predict tank crew performance have been performed at the Army Research Institute Field Unit, Fort Knox (Eaton, 1978; Eaton, Bessemer, & Kristiansen, 1979; Eaton, Johnson, & Black, 1980; Black, 1980); these studies were described in the earlier discussion of task performance criteria. Relationships between ASVAB and performance were generally low. In the few instances where significant relationships were found, they failed to cross-validate to new samples. At the Center for Naval Analysis, relationships between ASVAB scores and measures of performance of Marine Corps recruits were examined in an interim/progress report (Hiatt & Sims, 1980) to determine the feasibility of validating enlistment standards against job performance. Graphs present relationships of AFQT and educational level with several probabilities: completion of first-term enlistment (attrition), recommendation for reenlistment, promotion to corporal, and a composite of completion of term and promotion to corporal. Relationships are similar to those frequently reported for suitability criteria. Probability of promotion is reported as a reflection of time required to learn a job, an interpretation that some readers may question. Uncorrected median validity coefficients for AFQT and supervisor ratings in six MOSs range from about .01 to .24. Validities for ASVAB composites range from about .06 to .26. Project plans call for subsequent validation of ASVAB scores against job sample tests in three Marine Corps MOSs. Performance tests are now being constructed for this purpose at NAVPERS-RANDCEN.¹⁰

¹⁰Edward Pickering, NAVPERSRANDCEN. Personal communication.

In a study for the Army Research Institute, Maier and Grafton (1981) obtained the following composite validities for ASVAB Forms 8, 9, and 10 with total Army skill qualification test (SQT) score: combat .56, field artillery .63, electronics .59, general maintenance .73, and food service .61. The SQT generally consists of three components: the skills component (SC), a multiple-choice written test of job information; the hands-on-test (HOT), a performance test administered under standardized conditions; and the job site component (JSC), a supervisor-scored checklist of work on the job. The validities reported by Maier and Grafton are similar to those reported earlier for aptitude composite and job knowledge tests, and they suggest that the written component is making the major contribution to total score variance.

Total SQT score was also one of several criterion variables examined in a study that estimated the effects of the calibration error in ASVAB Forms 5, 6, and 7 (Greenberg, 1980). Soldiers who would have been ineligible for Army service had the ASVAB been correctly normed were identified, and their performances were compared with those of soldiers who would have been eligible. Major findings about job performance (as distinct from training performance) were:

1. Aptitude was related to SQT total score, whereas level of education had little influence on it (most high school nongraduates who performed poorly were separated from the Army before taking the SQT).¹¹
2. First-term attrition was nearly twice as great for high school nongraduates as for graduates, whereas variation in attrition was slight as a consequence of differences in aptitude.
3. High school graduates and persons with higher aptitude scores were more likely to be promoted to grade E-5, whereas soldiers at all educational and aptitude levels had a high rate of achieving grade E-4 if they completed their first term.
4. Soldiers who would have been ineligible for Army service with the ASVAB correctly normed had a higher attrition rate and lower SQT scores.

Although aptitude was found to be related to SQT scores in both studies (Maier & Grafton, 1981; Greenberg, 1980), Greenberg found that it was not highly related to training performance (course completion) for the same MOS. Greenberg suggests these possible explanations:

1. Most courses were not particularly demanding.
2. Trainees had been prescreened and probably could cope with the subject matter.
3. Graduation from training is a crude measure of success in training.

The following factors, which were not mentioned by Greenberg, also may contribute to the finding that aptitude is more highly related to SQT performance than to training completion:

1. Low-aptitude soldiers may forget what they learn more rapidly.

¹¹It should be noted that aptitude standards for high school nongraduates are higher than those for high school graduates. As a consequence, educational level is inevitably confounded with aptitude.

2. Most skill and knowledge necessary for SQT performance may be acquired after completion of training. Low-aptitude soldiers either may receive duty assignments that limit their opportunity for learning or may fail to learn for other reasons in an operational environment.

At the Rand Corporation, David Armor is using SQT scores as one of several criterion measures (others are training performance and attrition rates) in the development of trade-off models of force structure.¹² The costs and performance levels associated with recruiting, training, and maintaining service personnel who possess different attributes (e.g., aptitude, educational level, sex) will be determined. Aptitude cut scores and other characteristics will be manipulated to determine optimal points for maximizing performance and minimizing costs. While the study began with the use of Army data (largely because SQT scores were available in addition to the more conventional measures of performance), it is expected that ultimately data from all services will be examined.

DISCUSSION AND CONCLUSIONS

Major Patterns of Use

The most frequently used predictors in the studies reviewed for this effort were aptitude variables, followed by level of education and other readily obtained biographic and demographic information. The most frequently used criterion measure was supervisor ratings, followed by composite measures of suitability, which often contained ratings. Performance was much less frequently measured by proficiency tests of job knowledge and still less by job sample tests.

Relationships Among Predictors and Criteria

As noted earlier, findings are constrained by the combinations of predictors and criteria that have been used. Information about criterion measures is based mostly on the aptitude variables used as predictors. Under this constraint, job knowledge and job sample test performance were predicted with median validity coefficients of .40 and .31 respectively. Composite measures of suitability were predicted with a median coefficient of .24, and global ratings of performance were least predictable, with an average coefficient of .15.

Current Predictive Validities

When the review was begun, it was hoped that the strength of relationships could be estimated for most combinations of variables. The small number of cases has prevented this kind of analysis in most cells. Where enough data were available for analysis of particular combinations, median validities for aptitude with job knowledge fell in the .30 to .50 range; and for training grade with job knowledge, in the .40 to .50 range. Median validities for aptitude with job sample tests were in the .10 to .35 range. Supervisor ratings were predicted by aptitude and biographic/demographic variables with low median values of about .12 to .17. Supervisor ratings of performance were predicted best by earlier performance in training, with a median validity of .23.

¹²David Armor, Rand Corporation. Personal communication.

Possible Improvement in Predictive Validities

In some instances, it seems possible to predict moderately well. Many of the validities are high enough to be of value in selection. Although the occasional appearance of a fairly strong relationship suggests that improvement is possible, two fundamental weaknesses in the current status of prediction should be mentioned. The first is a lack of attention to the content, level, and variability of performance in particular jobs and the relation of these factors to decisions about what is to be predicted and when prediction should occur.

Whether it is reasonable to expect to demonstrate high validities depends on many factors: variability across job holders, job difficulty, levels of performance at entry and after various lengths of time, the effective ceiling in job performance, and how soon and by what percentage of incumbents the ceiling is reached. These determine what criterion should be predicted and when. The abilities that contribute to the performance of a task may change as practice occurs (Fleishman & Fruchter, 1960; Fleishman & Hempel, 1956), and the requirements of performance in a job may change when a person becomes more experienced (Ghiselli, 1956a). Thus, the variables that predict initial learning and performance may not maintain the same relationship with later performance.

The second weakness in the current state of prediction is that, while aptitude and other variables often discriminate among incumbents during their early time on the job, these differences tend to wash out with experience (Brown & Vineberg, 1960; Siegel & Leahy, 1974; Vineberg & Taylor, 1972a). Glickman and Kipnis (1960) have suggested that supervisors are driven to differentiate among job holders on nontechnical factors because selection and training have combined to eliminate differences in technical ability. Christal (1979) has suggested that aptitude should not be expected to predict performance where selection and training have reduced differences in technical ability among incumbents. When differences in performance become minimal within the first year on the job--regardless of whether selection and training or experience is the leveler--the school may be the only place where differences in proficiency, and the rate at which it is acquired, will be evident.

Yet, this review revealed no systematic effort to take into account the job characteristics that are pertinent to prediction. Is the measurement of technical proficiency appropriate and informative for all jobs? If there is little variance in performance after some period of time, when should performance evaluation occur? Are there jobs for which it is important to capture differences among incumbents while they are still evident? These questions seem rarely to be addressed. It is not possible now to answer with confidence the question of how well predictions can be made. On the basis of available data, there may be little more room for prediction than has already been accounted for. Common sense argues against this conclusion.

Reliance on Global Ratings

Little would be gained by addressing these issues, however, without adequate measures of performance, where an over-reliance on global ratings has occurred. Problems of halo and contamination make ratings unsuitable for the specificity of measurement implicit in the evaluation of technical performance. "The well known phenomenon of halo which affects such [rating scale] items . . . prevents their being independent measures" (Bayroff, Haggerty, & Rundquist, 1954). Ratings require a degree of familiarity by raters with the work of persons they are called on to evaluate that is often not present or even possible (Wilson et al., 1954; Wiley, 1975a).

Job Sample Tests

For measuring technical proficiency, several writers have suggested the advantages of job sample tests (Mackie, 1967; Guion, 1976). Yet, the expense of such tests makes their general use appear impractical. The absence of a substantial body of data from the hands-on component of the Army's skill qualification test program testifies to the difficulty of obtaining meaningful, objective measurement even in a quasi-operational application.

Furthermore, jobs in which performance tests are essential for valid measurement are the exception (e.g., those with skilled components like multilimb coordination). Most behavior is mediated by information, which knowledge tests have the potential to measure adequately. When technical proficiency is a relevant aspect of performance, tests of job knowledge may provide the most objective, practical means for assessing it, despite their general dependence on verbal ability. Even though they have sometimes been found not to correlate well with performance tests or job experience, job knowledge tests can share considerable variance with both if derived from carefully developed job analysis data.

Job Knowledge Tests

No one can be sure that the effort to construct truly valid knowledge tests will be undertaken, but methodologies certainly can be provided. It seems likely that considerable benefit would accrue by expending as much effort on developing knowledge tests as has occasionally been spent on developing behaviorally anchored rating scales. In any event, the first task "is to determine what is to be predicted . . . Little improvement can be expected . . . simply by predicting a trivial aspect of performance" (Guion, 1976).

Supervisor's Global Ratings

It must be concluded that global ratings should be used as measures of overall suitability, not of technical proficiency. In jobs where social skill and response to situational requirements are the only attributes worth measuring, ratings, which provide a strong reflection of social relationships between supervisor and ratee, would appear appropriate. However, jobs that may have significant technical demands call for investigating the actual and potential variability of performance with measures sensitive to technical ability.

Two Promising Approaches for Future Predictor Validation Development

There is evidence that an effective strategy for predicting performance is to maximize the match between the behavior sampled by predictors and the sample of behavior to be predicted, as well as the match between the methods of measurement used in sampling each. Asher and Sciarrino (1974) have referred to this as a "point-to-point" strategy, stating that "the more features in common between the predictor and the criterion space, the higher the validity." For example, as suggested in Table 4, a sample of technical performance on the job may generally be predicted best by a sample of technical achievement in training. Evaluations by instructors during training made a major contribution to the prediction of later evaluations by supervisors on the job (see especially Flyer, 1963). Support for the "point-to-point" strategy also comes from the Army Research Institute, Fort Knox Field Unit (Eaton, Johnson, & Black, 1980), in which job sample predictors have been used to predict job sample criteria.

The literature review revealed that performance in training is currently the best predictor of both job proficiency (as measured by job knowledge) and job performance (as

measured by supervisor ratings). At least two approaches to prediction would appear to take some advantage both of the point-to-point strategy and of the efficiency of training performance as predictor: (1) the use of miniaturized training and assessment centers in which prospective trainees can be tested in a sample of work activity, and (2) increased focus on individualized, self-paced training as a predictor. If the time for training and observation in the center were kept brief, the former approach could be used for entry screening for military service (performance in self-paced training could not, of course, be used for entry screening).

The miniaturized training and assessment center appears feasible if it is used selectively to predict performance in particularly demanding jobs that require extended training, such as electronic maintenance, automated data processing, or sonar operator.¹⁹ The size of the training investment for such jobs might warrant the added costs of prediction and the requirements of the complex jobs may make performance in them less predictable with conventional measures. In less demanding jobs (e.g., cook, clerk, security guard), the basis for predicting a person's performance seems more readily specifiable. Here, a stable personality, a preference for indoor or outdoor activity, interpersonal-relation skills, minimum aptitude and physical strength requirements, and so on may well be sufficient for predicting future performance. These requirements are predicted reasonably well on the basis of information derived from biographical, interest, and attitudinal inventories, as well as from conventional aptitude tests.

The second area of promise mentioned above is prediction of performance in self-paced training. This mode of instruction is increasingly in vogue and opportunities to use self-paced school performance in prediction can be expected to increase. The particular virtue of using self-paced training for prediction is that it may give evidence of abilities required in job performance but not tapped in the traditional classroom. Success in self-paced training appears to depend, to a greater extent than in traditional training methods, on individual initiative, motivation, decision-making ability, and the ability to perceive and adjust to the demands of a somewhat unstructured situation. Of course, the measures used at present to document individual performance in self-paced courses have not been designed for use as predictors, and some refinement and development of measures would be needed.

RECOMMENDATIONS

1. Greater attention should be given to characteristics of jobs and people in jobs when test validation studies are conducted. For example, the predictability of a criterion depends on many factors other than just the predictor-criterion relationships. Some of these factors are variability of performance across job holders, job difficulty, performance levels at entry and after various lengths of time, the effective ceiling in job performance, and how soon and by what percentage of incumbents the ceiling is reached. Greater understanding of the predictability of job performance criteria will require systematic study of these previously neglected factors in conjunction with the predictor-criterion relationships.

¹⁸To the extent that high-aptitude persons were selected for screening in the training and assessment center, this suggestion represents a departure from the recent use of this method with low-aptitude persons (Siegel & Bergman, 1972; Siegel, Bergman, et al., 1973; Siegel & Leahy, 1974; Siegel & Wiesen, 1977; Cory, in press).

2. The use of miniaturized training and assessment centers warrants further evaluation, especially in predicting performance for demanding jobs in which the size of the training investment may warrant the added costs of prediction.
3. Relationships among predictor, training, and job performance variables must be better understood. In this context, there should be a focus in self-paced training on variables that can serve as supplemental predictors to entry tests.
4. Use of supervisors' ratings as the sole measure of job performance should be restricted to jobs for which motivation, social skill, and response to situational requirements are the only attributes worth measuring.

REFERENCES

- Allusisi, E. A. Performance measurement technology: Issues and answers. In L. T. Pope & D. Meister (Eds.), Symposium proceedings. Productivity enhancement: Personnel performance assessment in Navy systems--October 12-14, 1977. San Diego: Navy Personnel Research and Development Center, undated.
- Asher, J. J. The biographical item: Can it be improved? Personnel Psychology, 1972, 25, 251-269.
- Asher, J. J., & Sciarrino, J. A. Realistic work sample tests: A review. Personnel Psychology, 1974, 27, 519-533.
- Austin, J. D. The relative efficiency of several predictors of airman proficiency tests scores (PRL-TM-55). Lackland Air Force Base, TX: Personnel Research Laboratory, Wright Air Development Division, January 1955.
- Bayroff, A. G., Haggerty, H. R., & Rundquist, E. A. Validity of ratings as related to rating techniques and conditions. Personnel Psychology, 1954, 7, 93-114.
- Black, B. A. ASVAB aptitude area score, CO, as a predictor of tank crewmember performance (Working Paper 80-9). Fort Knox, KY: Army Research Institute for the Behavioral and Social Sciences, October 1980.
- Brogden, H. E. On the interpretation of the correlation coefficient as measure of predictive efficiency. Journal of Educational Psychology, 1946, 37, 65-76.
- Brokaw, L. D. Prediction of Air Force training and proficiency criteria from armed forces selection tests (WADC-TN-59-194). Lackland Air Force Base, TX: Wright Air Development Center, Air Research and Development Command, October 1959. (a)
- Brokaw, L. D. Prediction of Air Force training and proficiency criteria from Airman Classification Battery AC-2A (WADC-TN-59-196). Lackland Air Force Base, TX: Wright Air Development Center, Air Research and Development Command, October 1959. (b)
- Brokaw, L. D. Prediction of criteria for medical and dental specialties from Airman Classification Battery AC-2A (WADC-TN-59-202). Lackland Air Force Base, TX: Wright Air Development Center, Air Research and Development Command, December 1959. (c)
- Brokaw, L. D. Suggested composition of airman classification instruments (WADD-TN-60-214). Lackland Air Force Base, TX: Wright Air Development Division, Air Research and Development Command, August 1960.
- Brown, G. H., & Vineberg, R. A follow-up study of experimentally and conventionally trained field radio repairmen (HumRRO Tech. Rep. 65). Alexandria, VA: Human Resources Research Organization, September 1960.
- Christal, R. E. What is the value of aptitude tests? In C. J. Mullins & W. R. Winn (Eds.), Criterion development for job performance evaluation: proceedings from symposium 23 and 24, June 1977 (AFHRL-TR-78-83). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, February 1979.

Cory, C. H. A comparison of the job performance and attitudes of category IVs and I-IIIIs in 16 Navy ratings (NPRDC Tech. Rep. 76-35). San Diego: Navy Personnel Research and Development Center, May 1976. (a) (AD-A024 642)

Cory, C. H. The predictive validity of operational and experimental variables for Mental Group IVs in the Navy: A review and summary of the findings from four sets of "culture fair" tests. Proceedings of the 18th Annual Conference of the Military Testing Association, 1976, 812-835. (b)

Cory, C. H. The assignment of general detail personnel in the Navy: Fleet follow-up of personnel appraised in a technical classification assessment center (NPRDC Tech. Rep.) San Diego: Navy Personnel Research and Development Center, in press.

Cory, C. H., Neffson, N. E., & Rimland, B. Validity of a battery of experimental tests in predicting performance of Navy Project 100,000 personnel (NPRDC Tech. Rep. 80-35). San Diego: Navy Personnel Research and Development Center, September 1980. (AD-A091 243)

Crowder, N. A., Morrison, E. J., & Demaree, R. G. Proficiency of Q-24 radar mechanics: VI. Analysis of intercorrelations of measures (AFPTRC-TR-54-127). San Antonio, TX: Air Force Personnel and Training Research Center, December 1954.

Curtis, E. W. Prediction of enlisted performance: I. Relationships among aptitude tests, Navy school grades, the report of enlisted performance evaluation, and advancement examinations (Tech. Bulletin STB 71-10). San Diego: Naval Personnel and Training Research Laboratory, June 1971.

Dunnette, M. D., & Borman, W. C. Personnel selection and classification systems. Annual Review of Psychology, 1979, 30, 477-525.

Dyer, F. N., & Hilligoss, R. E. Using an assessment center to predict filed leadership performance of Army officers and NCOs. Proceedings of the 19th Annual Military Testing Association (1977), 369-396.

Eaton, N. K. Predicting tank gunnery performance (RM 78-6). Fort Knox, KY: Army Research Institute for the Behavioral and Social Sciences, February 1978.

Eaton, N. K., Bessemer, D. W., & Kristiansen, D. M. Tank crew position assignment (Tech. Rep. 391). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences (PERI-IK), October 1979.

Eaton, N. K., Johnson, J., & Black, B. A. Job samples as tank gunnery performance predictors (ARIBSS Draft Tech. Rep. 447). Fort Knox, KY: Army Research Institute for the Behavioral and Social Sciences, November 1980.

Engel, J. D., & Rehder, R. J. A comparison of correlated-job and work sample measures for general vehicle repairman (HumRRO Tech. Rep. 70-16). Alexandria, VA: Human Resources Research Organization, October 1970.

Erwin, F. W., & Herring, J. W. The feasibility of the use of autobiographical information as predictor of early Army attrition (ARIBSS TR-77-A6). Washington, DC: Richardson, Bellows, Henry and Company, Inc., August 1977.

Fisher, W. E., Ward, J. H., & Holdrege, F. E. Prediction of unsuitability discharges (WADD-TN-60-260). Personnel Laboratory, Wright Air Development Division, October 1960.

Flanagan, J. C. The critical incident technique Psychological Bulletin, 1954, 51, 327-355.

Fleishman, E. A., and Fruchter, B. Factor structure and predictability of successive stages of learning Morse Code. Journal of Applied Psychology, 1960, 44, 96-101.

Fleishman, E. A., and Hempel, W. E. Factorial analysis of complex psychomotor performance and related skills. Journal of Applied Psychology, 1956, 40, 96-104.

Flyer, E. S. Educational level and Air Force adaptability criteria. Tri-Service Conference on Selection Research. Washington, DC: Office of Naval Research, 1960, 235-258. (a)

Flyer, E. S. Unreliable airmen in high-risk jobs: Unsuitability in the munitions and weapons maintenance career field (WADD-TN-60-43). Lackland Air Force Base, TX: Personnel Laboratory, Wright Air Development Division, March 1960. (b)

Flyer, E. S. Prediction of unsuitability among first-term airmen from aptitude indexes, high school reference data, and basic training evaluations (PRL-TDR-63-17). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, June 1963.

Flyer, E. S. Prediction by career field of first term airman performance from selection and basic training variables (PRL-TDR-64-5). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, March 1964.

Ghiselli, E. E. Dimensional problems of criteria. Journal of Applied Psychology, 1956, 40(1), 1-4. (a)

Ghiselli, E. E. Differentiation of individuals in terms of their predictability. Journal of Applied Psychology, 1956, 40(6), 374-377. (b)

Ghiselli, E. E. The validity of occupational aptitude tests. New York: Wiley, 1966.

Glickman, A. S., & Kipnis, D. Theoretical considerations in the development and use of a non-cognitive battery. Tri-Service Conference on Selection Research, 9-19. Washington, DC: Office of Naval Research, 1960.

Gordon, M. A., & Botenberg, R. A. Prediction of unfavorable discharge by separate educational levels (PRL-TDR-62-5). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, April 1962.

Gordon, M. A., & Flyer, E. S. Predicted success of low-aptitude airmen (PRL-TDR-62-14). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, August 1962.

Gragg, D. B., & Gordon, M. A. Validity of the Airman Classification Battery AC-1 (2nd ed.) (Research Bulletin 50-3). Lackland Air Force Base, TX: Human Resources Research Center, September 1951.

Greenberg, I. M. Mental standards for enlistment performance of Army personnel related to AFQT/ASVAB scores (MGA-0180-WRO-02). Washington, DC: McFann-Gray and Associates, Inc., December 1980.

- Grings, W. W., & Rigney, J. W. A methodological study of electronics troubleshooting skill: II. Intercomparisons of the MASTS test, a job sample test, and ten reference tests administered to Fleet Electronics Technicians. Los Angeles, CA: Department of Psychology, University of Southern California, 1953.
- Guinn, N., Kantor, J. E., Magness, P. J., & Leisey, S. A. Screening for entry into the security police career field (AFHRL-TR-77-79). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1977.
- Guion, R. M. Personnel selection. *Annual Review of Psychology*, 1967, 18, 191-216.
- Guion, R. M. Recruiting, selection, and job placement. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.
- Guion, R. M. Scoring of content domain samples: The problem of fairness. *Journal of Applied Psychology*, 1978, 63(4), 499-606.
- Harding, F. D., & Bottenberg, R. A. Contribution of status factors to relationships between airmen's attitudes and job performance (ASD-TN-61-147). Lackland Air Force Base, TX: Personnel Laboratory, Aeronautical Systems Command, November 1961.
- Hausman, H. J., & Strupp, H. H. Non-technical factors in supervisors' ratings of job performance. *Personnel Psychology*, 1955, 8(1), 201-217.
- Hiatt, C. M., & Sims, W. H. Armed Services Vocational Battery (ASVAB) and job performance (CNA 80-3121). Alexandria, VA: Center for Naval Analyses, Marine Corps Operations Analysis Group, November 1980.
- Hickerson, K. A., Hazel, J. T., & Ward, J. H., Jr. A causal analysis of relationships between performance and satisfaction in eight airman specialties (AFHRL-TR-75-57). Lackland Air Force Base, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, October 1975.
- Holberg, A., & Pugh, W. M. Predicting Navy effectiveness: Expectations motivation, personality, aptitude, and background variables (NHRC 77-53). San Diego, CA: Naval Health Research Center, November 1977.
- Judy, C. J. A regression analysis of one set of Airman Proficiency Test scores (WADD-TN-60-139). Lackland Air Force Base, TX: Wright Air Development Division, Air Research and Development Command, June 1960.
- Kipnis, D. Prediction of performance among lower aptitude men (Technical Bulletin 61-10). Washington, DC: Bureau of Naval Personnel, July 1961.
- Kipnis, D. A noncognitive correlate of performance among lower aptitude men. *Journal of Applied Psychology*, 1962, 46(1), 76-80.
- Kipnis, D. The relationship between persistence, insolence, and performance, as a function of general ability. *Educational and Psychological Measurement*, 1965, XXV (1), 95-110.
- Klieger, W. A. Dubuisson, A. U., & de Jung, J. E. Prediction of unacceptable performance in the Army (HFRB Tech. Res. Note, No. 113). Washington, DC: Human Factors Research Branch, TAG R&D Command, DA, June 1961.
- Landy, F. J., & Farr, J. L. Performance rating. *Psychological Bulletin*, 1980, 87(1), 72-107.

Lockman, R. F. Enlisted selection strategies (CNS 1039). Arlington, VA: Center for Naval Analyses, September 1974.

Mackie, R. R. Factors influencing the use of the practical performance tests in the U.S. Navy. Proceedings of the 9th Annual Conference of the Military Testing Association (1967), 32-40.

Mackie, R. R., & High, W. S. Research on the development of shipboard performance measures: Supervisory ratings and practical performance tests as complementary criteria of shipboard performance (Tech. Rep. IX). Los Angeles, CA: Human Factors Research, Inc., June 1959.

Mackie, R. R., Ridhalgh, R. R., Seltzer, M. L., & Shultz, T. E. Relationship between the armed services vocational aptitude battery (ASVAB) and surface sonar technician performance (NPRDC Tech. Rep. 81-19). San Diego: Navy Personnel Research and Development Center, August 1981. (AD-B059 929)

Mackie, R. R., Ridhalgh, R. R., & Shultz, T. E. New criteria for the selection and evaluation of sonar technicians (NPRDC Tech. Rep. 81-13). San Diego: Navy Personnel Research and Development Center, July 1981. (AD-B059 973L)

Mackie, R. R., Wilson, G. L., & Buckner, D. N. Research on the development of shipboard performance measures: Interrelationships between aptitude test scores, performance in submarine school, and subsequent performance in submarines as determined by ratings and tests (Tech. Rep. V). Los Angeles, CA: Management and Marketing Research Corporation, October 1954.

Maier, M. H. & Grafton, F. C. Aptitude composites for ASVAB 8, 9, & 10 (RR 1308). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences, May 1981.

Merenda, P. F. Navy Petty Officer promotion examinations as predictors of on-the-job performance. Educational and Psychological Measurement, 1959, XIX (4), 657-661.

Morsh, J. E. The development of Air Force Reference Battery II (AFPTRC-TN-57-104). Lackland Air Force Base, TX: Air Force Personnel Training Research Center, July 1957.

Mullins, C. J., & Winn, W. R. (Eds.) Criterion development for job performance evaluation: Proceedings from Symposium 23 and 24, June 1977 (AFHRL-TR-78-85). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, February 1979.

Plag, J. A. Pre-enlistment variables related to the performance and adjustment of Navy recruits. Journal of Clinical Psychology, 1962, 18, 168-171.

Plag, J. A., Goffman, J. M., & Phelan, J. D. The adaptation of naval enlistees scoring in mental group IV on the Armed Forces Qualification Test (NMNRU Rep. No. 68-23). San Diego: Navy Medical Neuropsychiatric Research Unit, 1967.

Plag, J. A., Goffman, J. M., & Phelan, J. D. Predicting the effectiveness of new mental standards enlistedees in the U.S. Marine Corps (NMNRU 71-42). San Diego: Navy Medical Neuropsychiatric Research Unit, 1970.

Plag, J. A., & Hardacre, L. E. The validity of age, education, and GCT score as predictors of two-year attrition among naval enlistees (NMNRU 64-15). San Diego: Navy Medical Neuropsychiatric Research Unit, June 1964.

Plag, J. A., & Hardacre, L. E. Age, years of schooling, and intelligence as predictors of military effectiveness for naval enlistees (NMNRU 65-19). San Diego: Navy Medical Neuropsychiatric Research Unit, July 1965.

Pope, L. T., & Meister, D. (Eds.). Symposium proceedings. Productivity enhancement: Personnel performance assessment in Navy systems--October 12-14, 1977. San Diego: Navy Personnel Research and Development Center, undated.

Saunders, D. R. Moderator variables in prediction. Educational and Psychological Measurement, 1956, 16 (1), 209-222.

Schmidt, F. L., & Kaplan, L. B. Composite vs. multiple criteria: A review and resolution of the controversy. Personnel Psychology, 1971, 24, 419-434.

Schultz, D. G., & Siegel, A. I. Post-training performance criterion development and application: A selection review of methods for measuring individual differences in on-the-job performance. Wayne, PA: Applied Psychological Services, July 1961.

Seashore, S. E., Indik, B. P., & Georgopoulos, B. S. Relationships among criteria of job performance. Journal of Applied Psychology, 1960, 44, 195-202.

Severin, D. The predictability of various kinds of criteria. Personnel Psychology, 1952, 5 (2), 93-104.

Shirkey, E. C. Preliminary validity report of the MOS evaluation test for medical specialist (Research Study No. 31). Indianapolis, IN: U.S. Army Enlisted Evaluation Center, Fort Benjamin Harrison, February 1965.

Shirkey, E. C. Preliminary validity report of the MOS evaluation test for general supply specialist (Research Study No. 38). Indianapolis, IN: U.S. Army Enlisted Evaluation Center, Fort Benjarnin Harrison, March 1966.

Siegel, A. I., & Bergman, B. A. Nonverbal and culture fair performance prediction procedures. I. Background, test development and initial results. Wayne, PA: Applied Psychological Services, Inc., June 1972.

Siegel, A. I., Bergman, B. A., & Lambert, J. Nonverbal and culture fair performance prediction procedures. II. Initial validation. Wayne, PA: Applied Psychological Services, Inc., September 1973.

Siegel, A. I., & Leahy, W. R. Nonverbal and culture fair performance prediction procedures. III. Cross validation. Wayne, PA: Applied Psychological Services, Inc., March 1974.

Siegel, A. I., & Wiesen, J. P. Experimental precedures for the classification of naval personnel (NPRDC Tech. Rep. 77-3). San Diego: Navy Personnel Research and Development Center, January 1977. (AD-A035 744)

Smith, P. C. Behaviors, results, and organizational effectiveness: The problem of criteria. In Dunnette, M. D. (Ed.), Handbook of industrial and organizatin psychology. Chicago: Rand McNally College Publishing Company, 1976.

Taylor, E. K., & Nevis, E. C. Personnel selection. Annual Review of Psychology, 1961, 12, 389-412.

- Taylor, H. C., & Russell, J. T. The relationship of validity coefficients to the practical validity of tests in selection: Discussion and tables. Journal of Applied Psychology, 1939, 23, 565-578.
- Thorndike, R. L. Personnel selection: Test and measurement techniques. New York: John Wiley and Sons, Inc., 1949.
- Tupes, E. C., Brokaw, L. D., & Kaplan, M. N. An application of the hierarchical factor model to the criterion grouping problem (WADD-TN-60-219). Lackland Air Force Base, TX: Personnel Laboratory, Wright Air Development Division, September 1960.
- Tuttle, T. C., & Hazel, J. T. Review and implications of job satisfaction and work motivation theories for Air Force research (AFHRL-TR-73-56). Lackland Air Force Base, TX: Occupational Research Division, Air Force Human Resources Laboratory, January 1974.
- Urry, V. W., Shirkey, E. C., & Waldkoetter, R. O. Preliminary validity report of the MOS Evaluation Test for Cook (Research Study No. 20). Indianapolis, IN: U. S. Army Enlisted Evaluation Center, Fort Benjamin Harrison, February 1965.
- Vineberg, R., & Joyner, J. N. Performance of men in different mental categories: 2. Assessment of performance in selected Navy jobs (Tech. Rep. 78-1). Alexandria, VA: Human Resources Research Organization, September 1978.
- Vineberg, R., & Taylor, E. N. Performance in four Army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance (HumRRO Tech. Rep. 72-22). Alexandria, VA: Human Resources Research Organization, August 1972. (a)
- Vineberg, R., & Taylor, E. N. Performance in four Army jobs by men at different aptitude levels: 4. Relationships between performance criteria (HumRRO Tech. Rep. 72-23). Alexandria, VA: Human Resources Research Organization, August 1972. (b)
- Wherry, R. J. Control of bias in rating: Survey of the literature (Tech. Rep. DA-49-0853 OSA 69). Washington, DC: Department of the Army, Personnel Research Section, September 1950.
- Wiley, L. N. Relation of job qualification ratings to performance ratings of basic training instructors (PRL-TDR-64-21). Lackland Air Force Base TX: Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, July 1964.
- Wiley, L. N. Describing airman performance in the administrative career ladder by identifying patterns of trait ratings (PRL-TR-66-13). Lackland Air Force Base, TX: Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, November 1966.
- Wiley, L. N. Across-time prediction of the performance of airman administrators and mechanics (AFHLR-TR-74-53). Lackland Air Force Base, TX: Occupational Research Division, Air Force Human Resources Laboratory, July 1974.
- Wiley, L. N. Familiarity with subordinates' jobs: Immediate versus secondary supervisors (AFHRL-TR-75-7). Lackland Air Force Base, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, June 1975. (a)

Wiley, L. N. Ratings of first-term airmen on supervisory potential and technical competence in AFSCs 462XO and 812XO (AFHRL-TR-75-56). Lackland Air Force Base, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, October 1975. (b)

Wiley, L. N. Airman job performance estimated from task performance ratings (AFHRL-TR-76-64). Lackland Air Force Base, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, October 1976.

Wiley, L. N., & Cagwin, L. P. Comparing prediction of job performance ratings from trait ratings for aircraft mechanics and administrative airmen (AFHRL-TR-68-108). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, October 1968.

Wiley, L. N., & Hahn, C. P. Task level job performance criteria development (AFHRL-TR-77-75). Brooks Air Force Base, TX: Air Force Human Resource Laboratory, December 1977.

Wilson, C. L., & Mackie, R. R. Research on the development of shipboard performance measures: Part I. The use of practical performance tests in the measurement of shipboard performance of enlisted naval personnel. Los Angeles: Management and Marketing Research Corporation, November 1952.

Wilson, C. L., Mackie, R. R., & Buckner, D. N. Research on the development of shipboard performance measures: Part III. The use of performance check lists in the measurement of shipboard performance of enlisted naval personnel. Los Angeles: Management and Marketing Research Corporation, February 1954.

Wiskoff, M. F., Atwater, D. C., Houle, M., & Sinalko, H. W. Enlisted first-term-attrition: Literature review and discussion. Paper presented at The Technical Cooperation Program, Subcommittee on Nonatomic Military Research and Development, Montreal, Canada, June 1980.

Yellen, T. M. I. Preliminary validity report of the MOS evaluation test for truck vehicle mechanic, MOS Code 632.2 (63C20) (Research Study 49). Indianapolis, IN: U.S. Army Enlisted Evaluation Center, Fort Benjamin Harrison, 1966.

BIBLIOGRAPHY

- Lecznar, W. B., & Davydiuk, B. F. Airman classification test batteries: A summary (WADD-TN-60-135). Lackland Air Force Base, TX: Wright Air Development Division, Air Research and Development Command, May 1960.
- Lent, R. H., Aurbach, H. A., & Levin, L. S. Research design and validity assessment. Personnel Psychology, 1971, 24, 247-274.
- Lent, R. H., Aurbach, H. A., & Levin, L. S. Predictors, criteria, and significant results. Personnel Psychology, 1971, 24, 519-533.
- Weeks, J. L., Mullins, C. J., & Vitola, B. M. Airman classification batteries from 1948 to 1975: A review and evaluation (AFHRL-TR-75-78). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1975.

APPENDIX
ABSTRACTS OF PUBLISHED STUDIES

1. Abrahams, N. M., Neumann, I., & Rimland, B. Preliminary validation of an interest inventory for selection of Navy recruiters (NPTRL Research Memorandum SRM 73-3). San Diego, CA: Naval Personnel and Training Research Laboratory, April 1973.

The quality and quantity of Navy enlisted personnel are in large part dependent upon the effectiveness of Navy recruiters. The advent of the all-volunteer armed forces has made selection of the most capable recruiters increasingly important.

The Strong Vocational Interest Blank (SVIB) has been used successfully by the Naval Personnel and Training Research Laboratory (NPTRL) to identify: (1) those individuals most likely to complete an officer training program such as the Naval Academy or NROTC, and (2) those individuals most likely to pursue a full Navy career. This report presents the preliminary findings of a research program aimed at improving recruiter selection through the use of the SVIB and other predictor instruments.

SVIBs were collected from samples representing the most and the least effective recruiters at 36 of the 42 main recruiting stations. The responses of the two groups were contrasted for one-half of the sample and used to establish scoring weights. The valid responses were assembled into the Recruiter Interest Scale-1 (RIS-1). The remaining recruiters, not used in scale development, were scored on the RIS-1 to determine how well the scale discriminates between the most and least effective recruiters.

An empirical SVIB scale, RIS-1, was found to discriminate quite well between the most and least effective recruiters. When scores of the "holdout group" were ordered and divided into fourths, the top quarter contained about three times as many effective recruiters as did the bottom fourth. It is therefore recommended that the RIS-1 scale be used to identify potentially effective recruiters among those volunteering for recruiting duty.

Several suggestions intended to increase the number of applicants for recruiting duty, including a Shipmate Nomination System, were proposed.

Efforts toward improving recruiter selection, involving the SVIB, other instruments, and criterion refinement, are continuing.

2. Alf, E. F., & Gordon, L. V. A Fleet validation of selection tests for Underwater Demolition Team testing (Bureau of Naval Personnel Technical Bulletin 57-6). San Diego, CA: Naval Personnel Research Field Activity, July 1957.

In a previous study, a battery of predictor tests was administered to 140 students entering Underwater Demolition Team (UDT) training. Of the

entering group, 64 were graduated into fleet teams. Approximately 15 months after the last of this group graduated, the present follow-up study was performed to determine the relationship between this predictor battery and fleet success.

Forced rankings were obtained for 50 of the original 64 graduates on a number of traits important for fleet success. Correlations were obtained between scores on the original predictor battery and forced rankings on "overall operating ability." Swimming scores were correlated with rankings on "swimming ability." Other traits were too highly correlated with the first criterion to warrant separate analysis.

Basic Test Battery (BTB) scores were significantly correlated with fleet success, while swimming and physical fitness measures were not. Two personality traits, Objectivity and Masculinity, had significant validities against this fleet criterion. Swimming test scores correlated significantly with rankings of swimming ability in the fleet.

The study concluded that swimming ability and physical fitness are important as predictors of UDT training success but not of fleet success. Cognitive measures (BTB), while unpredictable of UDT training success, predict fleet success. Therefore, both types of measures should be used for screening in the initial UDT training program.

3. Anderson, A. V., & Rimland, B. Form 2 of the Sonar Pitch Memory Test: II. Validation of the test (Bureau of Naval Personnel Technical Bulletin 57-7). San Diago, CA: Naval Personnel Research Field Activity, July 1957.

This report describes the validation of Form 2 of the Sonar Pitch Memory Test (SPMT) and compares its effectiveness with that of Form 1, the longer test it was designed to replace.

Forms 1 and 2 of the SPMT were administered during the first six months of 1956 at all three Naval Training Centers to all recruits meeting the minimum Navy General Classification Test and Navy Arithmetic Test scores required of sonar technician trainees. The primary criteria consisted of scores on three administrations of the doppler discrimination test used to measure achievement in the doppler training programs at the Fleet Sonar Schools at Key West and San Diego. The Key West sample was 102 cases. The San Diego sample ranged in size from 169 to 25, depending on the variables involved.

The validity coefficients were corrected for incidental restriction in range in the case of SPMT Form 1, and for explicit restriction on Form 2. In general, Form 2 was more valid than Form 1 in predicting achievement in doppler training, the median validity coefficients being .60 and .53, respectively. Form 1, however, was somewhat more valid in predicting two criteria

of secondary relevance, final course grades and grades in the "operations" phase of the course. Forms 1 and 2 correlated .68 in an unselected sample.

Data were available at Key West for 86 men who had taken still another SPMT form, the 1950 revision. Not enough information was available to permit appropriate corrections for restriction in range.

Because of its higher validity in predicting doppler achievement, and because of its several administrative advantages over Form 1, SPMT Form 2 is recommended to be continued in use as the operational test for selecting sonar technician trainees at the Naval Training Centers and Naval Reserve units.

4. Anderson, A. V. Relationships among aptitude, school, and shipboard measures for Sonarmen: I. A preliminary study (Report No. 22). San Diego, CA: Naval Personnel Research Field Activity, November 1952.

Six months after students were graduated from the San Diego Fleet Sonar School, school officials sent a follow-up shipboard rating scale to the Commanding Officer of each ship to which the graduates were assigned. This study analyzed the relationships between rating scale items and selection and school measures.

Data from 121 rating scales and matching student record cards were analyzed by correlational and factor analysis methods. Results obtained were: correlations between selection test scores and school grades on one hand and shipboard ratings on the other were consistently low; General Classification Test scores were unrelated to shipboard performance ratings; the Arithmetic and Clerical Aptitude Tests were the only positive test predictors of the rating scale measures; the Mechanical Test had a consistent negative relationship to the shipboard measures. A scatter diagram of MECH scores plotted against rated sonar stack performance indicated that low mechanical aptitude does not detract from ability to operate the sonar stack. The factor analysis of the rating form indicated that the ability to understand relative movement problems is considered important to sonar success. The study revealed two auditory acuity factors, one at low frequencies and one at high. Hearing loss in the poorer ear at frequencies greater than 2048 cycles per second was not related to either school or shipboard performance. Other findings: Even though the Attack Teacher grades had very little variance, they were significantly related to some of the rating scale measures; school grades in general, with the exception of the grade in code, had very little variance.

It was recommended that this study's findings be verified in a second sample. More research is needed on sonar student selection techniques, particularly in the area of spatial relationships. Grading of Attack Teacher performance and other aspects of school training should be studied and

improved. The problem of whether or not men should be trained to perform both sonar operator and sonar maintenance duties warrants further investigation. An improved shipboard rating scale should be constructed.

5. Anderson, A. V. Relationships among aptitude, school, and shipboard measures for Sonarmen: II. A determination of the stability of the relationships. (Report No. 33, TR-4) San Diego, CA: Naval Personnel Research Unit, April 1953.

This study was a replication and extension of the investigation reported in the study above, "Relationships Among Aptitude, School, and Shipboard Measures for Sonarmen: I." Its primary purpose was to determine the stability of the findings of that study. A secondary purpose was to determine whether measures of auditory acuity for the poorer ear and better ear are comparable.

Data from an additional 123 shipboard ratings scales and matching student record cards were analyzed by correlational methods. The results were compared with those obtained in the previous study. In general, correlations obtained in the new study were not significantly different. Within the range of hearing loss considered in the two studies, there appears to be no relationship between auditory acuity at high frequencies and anti-submarine sonar technician performance. Attack Teacher grades seem to be consistent predictors of shipboard ratings. High intercorrelations among items in the shipboard rating scale occurred in both studies and appear to be a characteristic of the type of rating scale employed. The negative relationships (between the Mechanical Test scores and the shipboard measures) obtained in the previous study did not recur in this study. Auditory acuity scores for poorer and better ears do not differ to any appreciable extent in their relationships to other measures.

The effect of varying amounts of hearing loss on ability to perform the auditory tasks required of sonar technicians should be experimentally determined. No change in the Basic Test Battery or Pitch Memory selection standards should be made at present. When sufficient data have been accumulated on the Shipboard Rating Scale for Sonarmen developed by NAVPRU (Project P10), an analysis of the relationships of that scale to selection and school measures should be conducted.

6. Asher, J. J., & Sciarrino, J. A. Realistic work sample tests: A review. Personnel Psychology, 1974, 27, 519-533.

Realistic work sample tests that were miniature replicas of criterion tasks were classified as motor or verbal. The motor work sample

had subjects performing physical manipulations, such as tracing a complex electrical circuit or operating a sewing machine, while verbal work samples required individuals to cope with people-oriented or language-oriented problems.

The work sample tests selected from the literature were those especially designed to represent on-the-job criterion behavior in a specific situation rather than ready-made standardized tests unless the latter had an unmistakable surface relationship to the criterion.

The guiding hypothesis was a point-to-point theory that the more features in common between the predictor and criterion space, the higher the validity. In a previous review (Asher, 1972), it was found that historical information from the scorable application blank was data with a point-to-point relationship with the criterion and had the highest predictive power from a list of standard predictors including intelligence, personality, interest, perception, motor skill, and mechanical ability.

Complex work sample tests that were miniature replicas of specific criterion behavior should also have a point-to-point relationship with the criterion.

When job proficiency was the criterion, realistic motor work sample tests had the highest validity coefficients second only to biographical information. Verbal work sample tests were not as high as the motor, but they were in the top half of the predictors. When the criterion was success in training, verbal work sample tests were more powerful in predicting success in training than in forecasting job proficiency. Verbal work sample tests had substantially more significant validity coefficients than the motor when there was a training criterion.

The point-to-point theory does not preclude other possible explanations, such as the interaction hypothesis, the work-methods hypothesis, and the transfer-of-experience hypothesis.

7. Benton, A. L., & Bechtoldt, H. P. The Enlisted Personal Inventory (Part I) as a predictor of personal adjustment after recruit training (Bureau of Naval Personnel Technical Bulletin 55-6). Iowa City, IA: State University of Iowa, Department of Psychology, June 1955.

This report was primarily concerned with the ability of the Enlisted Personal Inventory (Part I) to predict the incidence of discharge for reasons of personal unsuitability after the successful completion of recruit training.

The medical and service records of 724 discharged men who had taken the Personal Inventory during the first week of recruit training were analyzed and the findings related to Inventory score. The study was instituted 19 months after the men had entered recruit training. On the basis of their records, the men were classified into five categories: I. Normal termination of

active duty; II. Discharged for personal unsuitability during recruit training; IIA. Discharged for personal unsuitability after completion of recruit training; III. Discharged for somatic disability during recruit training; IIIA. Discharged for somatic disability after completion of recruit training.

Mean Personal Inventory scores of the several groups were: Group I: 3.0; Group II. 6.3; Group IIA. 4.4; Group III. 5.1; Group IIIA. 3.1. The mean score of Group I was significantly lower than those of Groups II, IIA and III. The mean score of Group II was significantly higher than that of Group IIA. The mean score of Group III was significantly higher than that of Group IIIA. Estimates of the predictive efficiency of the Inventory, by analysis of the proportions of men in each group placing at or above each Inventory score, indicated that: (a) In consonance with the results of earlier studies, a fair degree of discrimination between the normal group and groups discharged during the period of recruit training is achieved; and, in contrast, (b) no discrimination of practical significance between the normal group and the groups discharged after the completion of recruit training is achieved.

It was concluded that the Enlisted Personal Inventory (Part I) is a fair predictor of present status but possesses little value as a prognostic instrument, when "prognosis" is defined as the prediction of personal maladjustment a relatively short time after the successful completion of recruit training. While the Inventory failed to provide a discrimination of practical value, it did discriminate in a statistical sense between "normal" enlisted men and those who were found to be personally inadequate after the completion of recruit training. This positive finding offers ground for the hope that an instrument particularly designed for the prognostic purpose might be more successful.

8. Berkhouse, R. G., Woods, I. A., & Sternberg, J. J. Measurement and Prediction of foreign language speaking ability (PRB Technical Research Report 1115). Washington, D. C.: Personnel Research Branch, TAGO, DA, April 1959.

The research effort on foreign language speaking ability testing can be summed up briefly as follows:

Since speaking ability can be predicted with a high degree of confidence by paper and pencil tests of language fluency, most personnel considered for language assignments can be effectively screened by paper and pencil tests. The appropriate Army Language Proficiency Test is very adequate for the purpose.

For most language jobs in the Army, there is not sufficient justification for an individual speaking ability test. But for a small number of special language assignments, where there is a high premium upon high level speaking ability, the AFLST can serve as a useful adjunct to paper

and pencil tests. Preliminary selection should first be made upon the basis of ALP test performance. Unless a person under consideration for special assignment has a rating of at least "good" on the ALP test, he or she should not ordinarily be subjected to individual testing procedures.

9. Birnbaum, A. B., Sharp, L. H., Armore, S. J., Sprunger, J. A., & Bolanovich, D. J. Prediction of success in ordnance jobs (PRB Technical Research Note 58). Washington, D. C.: Personnel Research Branch, TAGO, DA, October 1956.

The objective of these studies was to evaluate the effectiveness of composites of Army Classification Battery (ACB) tests for predicting success in jobs for which personnel were trained at Ordnance School. ACB scores were compared with ratings of the job success of 671 Ordnance Storage Specialists (MOS 763), Small Arms Repairmen (421), Light or Heavy Artillery Repairmen (422-3), Machinists (443), and Welders (442).

The best predictors of success in Ordnance Storage Specialist duties were composites involving the Army Clerical Speed Test--unbiased estimates of validity, corrected for restriction in range, of .30 to .35. Success in the remaining jobs was best predicted by the two-test composite of the Automotive Information Test with either the Army Clerical Speed Test, the Arithmetic Reasoning Test, or the Pattern Analysis Test - validity-generalization coefficients of .28 to .41.

10. Black, B. A. ASVAB Aptitude Area Score, CO, as a predictor of tank crewmember performance (FKFU Working Paper 80-9). Fort Knox, KY: Army Research Institute for the Behavioral and Social Sciences, October 1980.

This research effort involved the use of one of the ten Aptitude Area Scores, CO (combat), as a predictor of armor crewmember performance at the end of initial training and later on the job.

Performance criteria consisted of two hands-on skill tests--one for gunner/loaders and one for drivers. Each test was administered twice--immediately upon graduation and 4 to 8 months later in the unit of first assignment.

The relationship between the predictor (CO) and both the end-of-course test and the unit-administered test was assessed for 60 gunner/loaders and 27 drivers by computing Pearson Product Moment r values. A significant relationship was found between CO and the unit-administered gunner/loader tests. However, CO was not a predictor of performance on these same tests when administered upon completion of training. CO was not related to either end-of-course or unit-administered drivers tests.

The author notes that, in regard to selection/assignment practices, the criteria against which predictors are validated are extremely important, as are the amount and type of training given before measurement of those criteria.

11. Booth, R. F., McNally, M. S., & Berry, N. H. Predicting performance effectiveness in paramedical occupations. Personnel Psychology, 1978, 31, 581-593.

Aptitude, age at enlistment, years of schooling prior to service entry, and number of suspensions or expulsions from school were considered as potential predictors of effectiveness among 2,835 Hospital Corpsmen (HMs) and 848 Dental Technicians (DTs). Effective performance in these occupational groups was defined as completion of HM or DT training, remaining on the job for at least 2 years, and advancing in job responsibility beyond a minimum apprenticeship level during the 2-year post-training criterion period. The composite validity of these four variables was .46 for predicting HM effectiveness and .35 for predicting DT effectiveness; the cross-validities for these composites were .48 for HMs and .35 for DTs. Occupation-specific odds-for-effectiveness (OFEs) that provide a means for standardizing the use of age and school experience variables in evaluating an individual's chances for performing effectively in these paramedical jobs were generated from the regression equations developed in these analyses. The validities of these occupation-specific OFEs as predictors of performance effectiveness were not enhanced by considering either the minority status or sex of job candidates.

12. Borman, W. C., Toquam, J. L., & Rosse, R. L. An inventory battery to predict Navy and Marine Corps recruiter performance: Development and validation (NPRDC TR 79-17). Minneapolis, MN: Personnel Decisions Research Institute, May 1979.

The objective of this study was to develop paper-and-pencil predictors of Navy and Marine recruiter performance and evaluate their validity. Accordingly, several measures of personality, vocational interests, and background were prepared (or selected) and administered to a geographically representative sample totaling 329 Navy and 118 Marine Corps recruiters. Scores on the predictor battery's items and scales were correlated with performance scores developed from supervisory, peer, and self ratings and from producting data (i.e., numbers of recruits enlisted). Estimated cross-validities for predictor composites were significantly different from zero for four of the five performance criteria in the Navy sample. They ranged from .17 to .31. Corresponding validity estimates for the Marine Corps sample ranged from .22 to .38, ($p < .01$ for three criteria, $p < .05$ for two criteria).

Recommendations from the study included:

1. Examine the predictive validity of the predictor composites developed in this project.
2. Assess the potential fakability of the predictor composites.
3. Develop additional paper-and-pencil measures of constructs that this study suggests are valid indicators of Navy and Marine Corps recruiter success.

13. Bowser, S. E. Noncognitive factors as predictors of individual suitability for service in the U. S. Navy (NPRDC TR 74-13). San Diego, CA: Navy Personnel Research and Development Center, April 1974. (AD-780 438)

This study is a pilot utilizing non-cognitive data sources in the prediction of individual suitability for service in the U.S. Navy. A methodology was developed which enables a logical selection of subsets of categorical predictors to optimize the prediction of suitability for service. The results support the contention that non-cognitive data sources are important and useful in prediction of success in the U.S. Navy.

14. Boyd, K. N., & Jones, H. H. An analysis of factors related to desertion among FY 1968 and FY 1969 Army accessions (AFHRL-TR-73-63). Alexandria, VA: Manpower Development Division, Air Force Human Resources Laboratory, January 1973.

Desertion was investigated among Army accessions who entered the service at a time when entrance requirements were less restrictive than at present for some personnel. Several personal and demographic factors were found to distinguish deserters from non-deserters. Implications for personnel selection and management are discussed on the basis of anticipated desertion rates for those with predisposing backgrounds prior to service entry.

15. Brokaw, L. D. Prediction of Air Force training and proficiency criteria from Airman Classification Battery AC-2A (WADC-TN-59-196). Lackland Air Force Base, TX: Wright Air Development Center, Air Research and Development Command, October 1959.

This Note reports the validity of the Airman Classification Battery AC-2A during the first 14 months of its administration. Data are presented for 46 specialties for which both technical training and job proficiency criteria were available, in the form of Final School Grades and Airman Proficiency Test scores. Technical training validities are given for an additional 20 technical schools. The expectation of some reduction of general validity as a function of maximizing differentiating power was realized. Slightly greater drops in general validity than had been anticipated were found in the mechanical and administrative aptitude clusters, while the remainder of the battery showed validity comparing favorably with the preceding Battery AC-1B. The AC-2A Battery demonstrated itself to be an effective instrument for differential classification; interpretation of its validities are made in this frame of reference. Current Air Force policies require a different kind of instrument for most effective recruitment and placement of new airmen.

16. Brokaw, L. D. Prediction of Air Force training and proficiency criteria from Armed Forces selection tests (WADC-TN-59-194). Lackland Air Force Base, TX: Wright Air Development Center, Air Research and Development Command, August 1959.

Appropriateness of the Armed Forces Qualification Test for use in Air Force pre-enlistment screening is indicated by data showing the positive correlation of AFQT scores with final grades in technical training courses and with scores on Airman Proficiency Tests. There is nothing in the data to suggest that the test could be changed in a manner to improve its across-the-board prediction of success in Air Force specialties.

17. Brokaw, L. D. Prediction of criteria for medical and dental specialties from Airman Classification Battery AC-2A (WADC-TN-59-202). Lackland Air Force Base, TX: Wright Air Development Center, Air Research and Development Command, December 1959.

Validation of Airman Classification Battery AC-2A for five enlisted medical specialties and for apprentice dental specialist training revealed generally satisfactory predictive efficiency for the General Aptitude Index. Success in the Apprentice Medical Material Specialist Course (90631) was not well predicted by any measure of the battery, AFQT score, age, or education. Although the Electronics Aptitude Index seemed as valid as the General Aptitude Index for the specialties treated in this study, there was no basis for recommending a change in the selective aptitude index. The reader was reminded that these data were collected during a period of emphasis of differential classification of enlisted personnel. New policies of Air Force recruitment will permit application of measures to maximize validity in all career fields in future Air Force testing programs.

18. Brokaw, L. D. Suggested composition of Airman Classification instruments (WADD-TN-60-214). Lackland Air Force Base, TX: Wright Air Development Division, Air Research and Development Command, August 1960.

Each test of Airman Classification Battery AC-2A was evaluated for its contribution to Air Force classification procedures. Criteria were success in Air Force technical training and scores achieved on job proficiency tests. By a multiple regression technique standard, beta weights and a squared multiple correlation coefficient were derived for 16 predictors against both criteria for 36 criterion groups. Components for four aptitude indexes were selected by reviewing the frequency with which tests appeared among the best four predictors within each of four job clusters.

19. Brown, G. H., Wood, M. D., & Harris, J. D. Army recruiters: Criterion development and preliminary validation of a selection procedure (HumRRO FR-ED-75-8). Alexandria, VA: Human Resources Research Organization, April 1975.

Research in support of the Army's recruiting operations was conducted to (a) develop a valid criterion of recruiter effectiveness, and (b) develop and evaluate a recruiter selection test battery. Using data from a sample of 400 recruiters, statistical analyses were performed to determine the theoretical yield to be expected from each recruiter's territory based on a multiple correlation between territorial characteristics and production records. A formula was developed to express each recruiter's effectiveness, comparing his actual production with the predicted production. In Task B, tests were assembled to measure recruiter characteristics considered likely

to be associated with recruiting effectiveness: verbal fluency, sociability, achievement motivation, empathy, maturity/responsibility, and various background characteristics. The tests were administered to 45 highly successful, and to 43 very unsuccessful, recruiters. None of the individual test scores discriminated significantly between good and poor recruiters. One performance measure of verbal fluency did discriminate significantly, as did about 20 background-information items. The true value of these items for recruiter selection cannot be known until cross-validation has been accomplished.

20. Carleton, F. O., Burke, L. K., Klieger, W. A., & Drucker, A. J. Validation of the Army Personality Inventory against a military adjustment criterion (PRB Technical Research Note 71). Washington, D. C.: Personnel Research Branch, TAGO, DA, May 1957.

The Army Personality Inventory was constructed in 1947 to identify Army personnel not likely to make good soldiers because of adverse personality and behavior characteristics. Research conducted from 1947 to 1950, based on administration of the test to approximately 11,000 enlisted men in six Replacement Training Centers, led to the development of methods of scoring the test geared to the task of predicting which personnel would have favorable and which unfavorable tours of duty. The research, based on analysis of the scores made by 3,000 men discharged before 1949, gave evidence that the test was moderately successful for the above purpose.

This study was undertaken to improve earlier results with analysis of scores of 3,000 more men of the original group who were discharged between 1949 and 1954. New scoring methods were devised, this time against an improved system of identifying the favorable and unfavorable men. The new scoring methods did not improve upon the old methods for predicting favorable and unfavorable behavior in the Army, but did demonstrate the successful use of special scoring devices in increasing the predictive efficiency of personality tests like the API. These devices control the effect of faking or response distortion on the total score.

21. Corts, D. B. Validation of the commander's evaluation report and the MOS evaluation test for Field Radio Repairman MOS Code 31E40 (Technical Research Study 107). Indianapolis, IN: Army Enlisted Evaluation Center, December 1967.

The present validation report furnished pertinent validity data for the commander's evaluation report (CFR) and the MOS evaluation test for Field Radio Repairman, MOS Code 31E40, which were administered in the November 1966 evaluation period. The empirical validities were obtained for all predictors and combinations of predictors. The most appropriate utilization of the existing predictors was determined. Evaluative statistics were provided for CER appraisal and revision.

The following three methods of combining predictors were compared for the development of the composite scores (RCSs); (1) the present procedure of combining the total MOS evaluation test and the total CER as prescribed by Department of the Army directive; (2) the weighting of the total MOS evaluation test and the total CER by statistical procedures; and (3) the weighting of the subdivisions of the MOS evaluation test and CER by statistical procedures. The validities, after shrinkage, by method were: Method (1) .58; method (2) .57; and method (3) .60. Computational formulas were provided under Discussion for the development of RCSs by methods (2) and (3). Method (3) was found to be superior to the others. The total CER had a validity coefficient of .49 (significant at the .01 level) and the total MOS evaluation test had a validity coefficient of .48 (significant at the .01 level). Partial correlation coefficients for the total CER and the CER scales with the MOS evaluation test held as a constant indicated that 4 of the 12 scales were independently valid predictors of job performance.

22. Cory, C. H. The assignment of general detail personnel in the Navy: Fleet follow-up of personnel appraised in a Technical Classification Assessment Center (NPRDC TR). San Diego, CA: Navy Personnel Research and Development Center, in press.

Follow-up in the Fleet was carried out to validate scores from a Technical Classification Assessment Center for a small, exploratory sample of General Detail personnel. Criteria were supervisory ratings of on-job performance and two binary variables: retention/attrition and Striker/non-Striker status. The assignment recommendations of the Assessment Center for personnel in the study were found not to have been followed by Fleet commands. However, scores of the Assessment Center usefully supplemented the ASVAB classification tests and biographical variables as predictors of supervisory performance ratings. Further development validation of Assessment Center variables on a larger, more definitive sample was recommended.

23. Cory, C. H. A comparison of the job performance and attitudes of Category IVs and I-IIIs in 16 Navy ratings (NPRDC TR 76-35). San Diego, CA: Navy Personnel Research and Development Center, May 1976. (AD-A024 642)

As an aid to the appropriate assignment of Category IV personnel to Navy ratings, this study was intended to provide objective data on the performance abilities of IVs in a representative sample of ratings. Supervisory evaluations, biographical and information, and attitude data were collected on samples of IV and non-IV personnel in 16 Navy enlisted ratings. Comparisons of IVs and non-IVs in each rating were made in terms of job performance, personal characteristics, and attitudes. t tests were used to identify the distinguishing characteristics of high performing IVs in five ratings. Multiple-regression analyses were used to investigate the predictability of performance of Category IVs in three ratings.

In the ratings covered, IVs exhibited generally widespread but small deficits in on-job performance when compared with non-IVs. Deficits in the

global performance of IVs were generally statistically significant for the Boiler Technician, Machinery Repairman, and Quartermaster-Signalman ratings and rating groups. Test scores and educational attainment were associated with high on-job performance of IVs. There were few consistent differences in motivation and outlook between IVs and non-IVs.

24. Cory, C. H. An evaluation of computerized tests as predictors of job performance: II. Differential validity for global and job element criteria (NPRDC TR 76-28). San Diego, CA: Navy Personnel Research and Development Center, January 1976. (AD-A020 867)

This report, the second of two, presents data concerning the validity of a set of experimental computerized and paper-and-pencil tests for measures of on-job performance on global and job elements. It reports on the usefulness of 30 experimental and operational variables for predicting marks on 42 job elements and on a global criterion for Electrician's Mate, Personnelman, Sonar Technician, and Apprenticeship rating groups.

About 10 percent of the zero-order validities of experimental tests were statistically significant, with most of the significant validities being for the Sonar Technician rating. Most experimental tests with significant validities were computer-administered. Experimental variables substantially enhanced the predictive accuracy of the operational battery with the most useful increments being for the Sonar Technician rating.

There was little or no evidence of consistency of the job element characteristics across ratings. The job elements that were highly predictable were those that were important and central to the duties of particular ratings. For the Technical ratings, the most effective predictors of job element marks from the marks for job elements did not result in any practical increase in validity coefficients. Generally, low correlations were found between empirically-derived estimates of importance of personal attributes for particular job elements and similar estimates based on the judgments of personnel experts. Synthetic validity was generally not as accurate as multiple regression for predicting job performance.

25. Cory, C. H. The predictive validity of operational and experimental variables for Mental Group IVs in the Navy: A review and summary of the findings from four sets of "culture fair" tests. Proceedings of the 18th Annual Conference of the Military Testing Association, (1976) 812-815.

This report summarizes results of a four-phase study undertaken by the Navy to develop and validate personnel selection tests that would predict performance characteristics of low mental ability personnel.

Study findings suggested that, for Mental Group IVs who apply for enlistment in the Navy in the future, tests in the Navy Classification Battery together with measures of vocational interest, such as the Strong Vocational Interest Blank, can be used to select with considerable accuracy the IVs who have the most potential for advancing into Technical ratings.

26. Cory, C. H., Neffson, N. E., & Rimland, B. Validity of a battery of experimental tests in predicting performance of Navy Project 100,000 personnel (NPRDC TR 80-35). San Diego, CA: Navy Personnel Research and Development Center, September 1980. (AD-A091 243)

This report summarizes results of a four-phase study that originated as part of the Project 100,000 research effort. The purpose of the study was to develop "culture fair" aptitude tests that would permit the Navy to identify potentially successful recruits from those who scored low on conventional tests.

Nineteen experimental test-questionnaires were developed to measure practical (as opposed to academic) mental abilities. The experimental instruments were divided into four batteries, each of which was administered to a separate sample ranging in size from 5,000 to 12,000 recruits. The instruments were validated against supervisory performance ratings, rating progression, and retention criteria for sample members. Separate analyses were done for Mental Level IVs, Blacks, and for apprenticeship level (non-rated, undesignated strikers) and technical rating groups.

The findings were generally negative. With only a few exceptions, the experimental tests were not valid predictors of on-job performance for any of the subgroups studied, and were less valid than the conventional tests for predicting either job performance or rating advancement. Also, because of the wide variety of "culture fair" tests evaluated, it is unlikely that paper-and-pencil tests can be found that will identify previously overlooked aptitudes in low-ability populations. A number of by-product findings of potential value in optimizing the utilization of low aptitude personnel were provided from the analyses.

The findings were communicated to Navy and DoD officials when Project 100,000 was terminated; this report provides a record of these efforts to facilitate future research in the problem area.

27. Crowder, N. A., Morrison, E. J., & Demaree, R. G. Proficiency of Q-24 radar mechanics: VI. Analysis of intercorrelations of measures (AFPTRC-TR-54-127). San Antonio, TX: Air Force Personnel and Training Research Center, December 1954.

In this study, a large amount of data on the aptitudes, background, attitude, proficiency test scores, and job performance were collected from 155 flight-line mechanics assigned to maintenance of navigational and bombing equipment. This report contains the analysis of intercorrelations of the measures, which furnish a major part of the findings of the study.

Analyses were carried out to determine the predictive value of (1) proficiency tests in regards to supervisor rankings and peer ratings of job performance, and performance test scores, and (2) aptitude tests with respect to personnel selection, proficiency test scores, and supervisor rankings of job performance.

Results favored aptitude tests of inductive reasoning and multiple-choice proficiency tests for their predictive value in the selection of electronics maintenance personnel. With respect to methods of obtaining evaluations of on-the-job performance, the results showed peer ratings as having higher correlations with proficiency tests than supervisor's rankings.

28. Curtis, E. W. Prediction of enlisted performance: I. Relationship among aptitude tests, Navy school grades, the report of enlisted performance evaluation, and advancement examinations (Technical Bulletin STB 71-10). San Diego, CA: Naval Personnel and Training Research Laboratory, June 1971.

The tests of the Navy Basic Test Battery (BTB), used in classifying the 100,000 men entering the Navy yearly, are constructed, evaluated and employed largely to predict final grade in Class "A" schools. The purpose of the research reported was to analyze the psychometric and substantive characteristics of final grade, and to determine its relationships with subsequent performance and advancement examination scores. A secondary purpose was to ascertain the extent to which the breadth and usefulness of the BTB may have been delimited by the emphasis upon final grade as the criterion.

A special battery of 11 "factor pure" aptitude tests was administered to samples of beginning students at 19 selected Class "A" schools which prepare enlisted men for 11 Navy ratings. Each student's complete grade record, consisting of scores on all written and performance examinations, number of failing grades received, and final grade, was obtained at the end of the course. Approximately 2 years later, the official performance evaluations and advancement examination scores were obtained for about 3,000 of the 4,451 sample members. These data and the BTB scores of the sample members were analyzed using multiple-regression and correlation methods. Additional on-the-job performance data collected are being analyzed for presentation in a later report.

Reliability estimates for final grade indicated acceptable reliability for nine of the schools, and marginal reliability for the other ten schools. Subgrades pertaining to Morse Code and Teletype training were relatively unreliable.

The weights applied to subgrades in computing final grade usually reflected the statistical contributions the subgrades made toward final grade. However, some large, and many small, discrepancies were found.

The BTB was more adequate for predicting final grade than for predicting failure. In addition, the choice of selection tests for each school has not been optimum for predicting failure. Results of the 11 aptitude tests indicated that it would be possible to improve the BTB and reduce the fail-rate.

Total score on the Navy's on-the-job performance evaluation form correlated significantly with final grade in most of the schools, although many of the correlations were below .20. None of the aptitude tests showed much promise for predicting total score on the Navy form. Advancement examination scores correlated substantially with many of the aptitude tests, with many Class "A" school subgrades, and with final grade for most of the schools.

It is recommended that an adaptation of the test called Associative Memory be added to the BTB. Consideration should be given to evaluating adaptations of the "Numerical" and "Mechanical Information" tests during the next validation and revision of the BTB. An experimental study and evaluation of clerical aptitude tests is recommended in order to define and isolate the factor that predicts performance evaluations.

29. Doll, R. E., & Gunderson, E. K. E. Occupational group as a moderator of the job satisfaction-job performance relationship (NMNRU Report No. 69-11). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, April 1969.

Personnel comprising wintering-over parties at small scientific stations in Antarctica represent two broad but quite different occupational groups: civilian scientist and Navy enlisted men. The motivations of the

Navy enlisted men who volunteer are less related to their specific jobs in the Antarctic than are those of the civilian scientists. The results confirmed the hypothesis that occupational group is a moderator of the job satisfaction-job performance relationship, and that the relationship is more pronounced for the scientist group than for the Navy enlisted group.

30. Dyer, F. N., & Hilligoss, R. E. Using an assessment center to predict field leadership performance of Army officers and NCOs. Proceedings of the 19th Annual Military Testing Association (1977), 369-396.

The assessment center concept involves the immersion of individuals into situations that simulate those they would face if selected for entry or promotion. The concept has been widely used in industry and business to select personnel for high level positions. In 1973-1974, the U.S. Army Infantry School (USAIS) Assessment Center (ACTR) assessed students from the Infantry Officer Advanced Course (IOAC), the Infantry Officer Basic Course (IOBC), and the Advanced NCO Educational System (ANCOES) to determine the feasibility of the assessment center concept as a leadership development and leadership prediction technique. It also assessed students from the Branch Immaterial Officer Candidate Course (BIOCC) to determine the feasibility of the assessment center concept as a selection device. This paper discussed the effectiveness of the ACTR for predicting field leadership performance.

Self-Description Instruments provided the largest proportion of criterion predictors and also provided these scores with the least assessor and assessee time. On the other hand, the most assessor-intensive formal ACTR exercises actually did the poorest job of predicting the field leadership criterion. Intermediate between these extremes is the Entry Interview, which provided a fair number of predictors with only a moderate amount of assessor and assessee time.

31. Eaton, N. K., Johnson, J., & Black, B. A. Job samples as tank gunnery performance predictors (ARISS Draft TR 447). Fort Knox, KY: Army Research Institute for the Behavioral and Social Sciences, November 1980.

This research was designed to develop and evaluate job samples as predictors of tank gunnery performance. It was conducted in three phases. In Phase I, three job samples, representing three major requirements of tank gunnery performance, were developed and evaluated. These were (1) the requirement to properly track a target, (2) the requirement to sense the location of a fired round with respect to the target, and (3) the requirement to properly adjust the second round after a first-round miss. Each of these was tested with an appropriate simulator, yielding relatively objective performance measures. The criterion used to evaluate the proposed predictors consisted of a modified Table VI live-fire gunnery exercise.

Phase II research was designed to complement and expand upon Phase I, using a larger sample. In Phase I, the task measures were obtained from research participants who were completing training as tank gunner/loaders. To determine whether the relationships observed were a function of achievement or aptitude, Phase II research included 10 drivers who had recently completed (MOS 19F) driver training at Fort Knox, but had not been given extensive gunnery training.

In Phase III, the effect of two key variables, verbal feedback and level of prior training, on job sample-tank gunnery relationships were evaluated. In addition, a new job sample, center-of-mass, was included in the evaluation. Research participants were 31 individuals from the Reception Station at Fort Knox and 57 individuals in their eighth week of Basic Armor Training (BAT). Difference scores, a reflection of the amount of improvement over trials on-the-job sample tasks were evaluated as predictors of live-fire gunnery performance. Gunnery performance was scored using video playback techniques.

The results from the three phases of research suggest that job samples seem to offer promise in predicting performance after formal training but prior to assignment to operational units. Future research efforts may be directed toward the use of job samples as performance predictors for personnel within operational units. Hands-on/job sample tasks may be developed which are useful in the selection of gunners and tank commanders to fill vacated slots in operational units.

32. Eaton, N. K. Predicting tank gunnery performance (RM 78-6). Fort Knox, KY: Army Research Institute for the Behavioral and Social Sciences, February 1978.

There were four primary objectives in this study, to determine the relationships between tank commander's (TC) and gunner's (G) gunnery performance and aptitude test scores, skills test scores, and aptitude composite scores, and to determine the relationship between driver's (D) aptitude test scores and driver performance as measured by driver's rankings within their platoon.

Data were collected on 51 TCs, Gs, and Ds in a TOE Armor Battalion undergoing annual tank gunnery training and qualification. Paper and pencil aptitude instruments and performance and skills tests were used to predict tank gunnery performance and driver rankings.

Results suggested that 6 of the 9 aptitude tests administered (object completion, visual recognition, lateral perception, attention to detail, mechanical aptitudes, speed of perception) and 2 of 87 skills tests (gun-laying time, Willey BOT time) had potential for tank gunnery performance prediction.

33. Eaton, N. K., & Johnson, J. R. Prediction of tank gunnery using job samples. Proceedings of 21st Annual Conference of the Military Testing Association (1979), 670-678.

This research effort was conducted in two phases. The purpose of the Phase I study was to evaluate the relationship between performance on three job samples (tracking, sensing, and round adjustment) and tank gunnery performance. The results of the Phase I research revealed significant relationships between gunnery performance and both round sensing and tracking of a diamond figure. The fewer errors research participants made in sensing and tracking, the better they performed in tank gunnery.

Phase II was designed to complement and expand upon the Phase I research. The results confirmed both of the significant relationships between tank gunnery scores and diamond and sensing error. In addition, the relationships between gunnery and job sample scores seem to be more likely due to aptitude rather than achievement measurement. This is because gunner/loaders who had considerable gunnery training, scored no better on the job sample tasks than drivers, who had relatively little gunnery training.

Overall, it appears that the development and validation of an appropriate set of job samples gives promise of measures yielding reasonably large correlations with gunnery performance, and which have potential for use in assignment of personnel to appropriate training programs.

34. Eaton, N. K., Bassemer, D. W., & Kristiansen, D. M. Tank crew position assignment (Technical Report 391). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences (PERI-IK), October 1979.

This research was conducted to determine whether available paper-and-pencil aptitude and training measures could be used to predict tank driver, gunner, and tank commander performance, and if so, to develop appropriate prediction equations based on the aptitude measures.

The research was conducted in three phases. The first two phases were conducted with armor trainees at Fort Knox, and dealt with the gunner and driver positions. The third phase was conducted with armor crewmen in operational armor battalions, and dealt with the tank commander and gunner positions. In Phases I and II, at Fort Knox, measures of trainee aptitudes, training performance, driving performance, and main-gun tank gunnery were collected for trainees in the sample. Aptitude measures included the Armed Services Vocational Aptitude Battery (ASVAB) raw scores and additional paper-and-pencil tests, while training measures included performance on tests relating to tank weapons, maintenance, communication, etc. The criterion performances were tank commander ratings of trainee M60 tank driving on a standardized course and number of hits during main-gun tank firing. During Phase III, aptitude and main-gun firing measures were collected for tank commanders and gunners in a sample from a USAREUR armor division. Aptitude measures were based on a battery of paper-and-pencil tests. Gunnery measures were based on performance during tank crew qualification firing at Grafenwoehr, West Germany.

With armor trainees at Fort Knox, a number of potentially useful predictor variables were identified in Phase I. These included four ASVAB tests and three additional paper-and-pencil tests as gunnery predictors, and six ASVAB tests and two additional paper-and-pencil tests as driving predictors. Only one of the driving predictor tests was validated in Phase II, and none of the paper-and-pencil tests was correlated with the gunnery measure. Nevertheless, certain methodological problems entered Phase II, so the failure to validate the other tests did not necessarily indicate a true lack of relationship with criterion performance. In Phase III, conducted with operational units, none of the tank commanders' or gunners' paper-and-pencil test scores was correlated with tank crew qualification gunnery scores.

The results from Phases I and II suggest that the continuing need to make optimal assignments of Army recruits to gunner/loader or driver training may best be addressed by continued research on the paper-and-pencil measures identified in Phase I, as well as the exploration of other techniques such as job sample performance measurement. In continued research with the paper-and-pencil tests, formulas based on both regression-based models and unit-weighted models seem appropriate. The results from Phase III indicate that paper-and-pencil tests do not seem to offer promise of predicting performance of personnel in operational units on tank crew qualification gunnery. Perhaps research efforts could best be directed toward the development and empirical validation of job sample and simulator techniques based on sound task analyses. Such job sample/simulator research might also lead to measures to supplement prediction of gunnery performance for armor trainees.

35. Egbert, R. L., Meeland, T., Cline, V. B., Forgy, E. W., Spickler, M. W., & Brown, C. Fighter I: A study of effective and ineffective combat performers (HumRRO Special Report 13). Presidio of Monterey, CA: Army Leadership Human Research Unit, Human Resources Research Office, March 1958.

The identification of psychological characteristics of the good fighter as contrasted with the nonfighter is a necessary initial step in a long-range program concerned with optimum utilization of men in combat. Knowledge of these characteristics opens up the possibility of developing experimental procedures for selection, training, and organization of fighting units.

This research involved 310 men, identified as fighters or non-fighters from information supplied by their peers in Korean combat. Each of these subjects underwent extensive psychological testing. The major analyses dealt largely with the native-born white sample. The descriptions of the fighter and nonfighter indicate that the fighter tends to: be more intelligent; be more masculine; be a "doer"; be more socially mature; be preferred socially and in combat by his peers; have greater emotional stability, more leadership potential; have better health and vitality, a more stable home life, a greater fund of military knowledge; and have greater speed and accuracy in manual and physical performance. A previously issued report, HumRRO TR 44, deals primarily with the findings of this study. The present report emphasizes the methodology.

36. Erwin, F. W., & Herring, J. W. The feasibility of the use of autobiographical information as a predictor of early Army attrition (ARIBSS TR-77-A6). Washington, D. C.: Richardson, Bellows, Henry and Company, Inc., August 1977.

Experimental autobiographical questionnaires were administered to samples of incoming Army enlistees at Forts Dix and Jackson. Basic Combat Training and 180-day success and attrition data were collected, and questionnaires were item analyzed using attrition as a criterion. Subject to verification, scoring systems developed and score validities indicated that the use of standardized autobiographical questions would be reasonably successful in predicting 180-day Army attrition. Results were similar for both black and white subgroups.

37. Fisher, A. H., Jr. Army "New Standards" personnel: Relationships between literacy level and indices of military performance (HumRRO TR-71-6). Alexandria, VA: Human Resources Research Organization, April 1971. (also published as AFHLR-TR-71-12)

The Armed Forces have been accepting low mental level personnel under Project 100,000 since October 1966. Over 15% of these men read below the fifth-grade level at entry into service. This research was designed to determine the relationship between military performance and literacy status of a sample of these "New Standards" men after 23 months of service, and to develop an equation for predicting 23-month literacy status. Twenty-three month reading scores of approximately 3,000 Army men were dichotomized at the fifth-grade level, and the two groups compared on various indices of military performance. A regression equation was then developed for predicting literacy status on the basis of entry characteristics.

Literacy status at 23 months was found to be only slightly related to most of the performance and status indices. It is possible to predict 23-month literacy status reasonably well on the basis of information obtained at the time of entry into service.

38. Flyer, E. S. Educational level and Air Force adaptability criteria. In Tri-Service Conference on Selection Research, 235-58 (Office of Naval Research, Washington, D. C., 1960.)

The purpose of this report was (1) to describe some findings from a large-scale investigation of airmen unsuitability, particularly certain analyses involving educational level, and (2) to report results of a method

for obtaining ratings of preservice school behavior that may have considerable relevance for adaptability criteria.

One unique finding from the investigation was that educational level possesses so much predictive power when aptitude is held constant. Another finding concerns a questionnaire on preservice school behavior, constructed to compare high school graduates and nongraduates. Preliminary results indicated that, although many of the questionnaire items are highly correlated with AFQT and with high school graduation status, there may be considerable unique variance relevant for military performance criteria.

It is suggested that further research be done to determine the usefulness of questionnaire data of the type shown in this report in adding to high school graduation status and aptitude as predictors of service performance criteria.

39. Flyer, E. S. Factors relating to discharge for unsuitability among 1956 Airman accessions to the Air Force (WADC-TN-59-201). Lackland Air Force Base, TX: Personnel Laboratory, Wright Air Development Center, December 1959.

This report provides major findings from a large-scale research investigation in which suitable and unsuitable airmen were compared for a number of personal attributes. Educational level was found to be the best single predictor of unsuitability discharge, although aptitude and age considered in conjunction with educational level increased significantly the accuracy of prediction. The implications of the findings for current selection procedures are discussed.

40. Flyer, E. S. Prediction by career field of first-term airman performance from selection and basic training variables (PRR-TDR-64-5). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, March 1964.

To gain information that might be useful in improving airman classification, 29 predictor variables were evaluated by multiple regression techniques against a criterion of satisfactory performance during the first two years of enlistment. Variables included personal data, educational and aptitude data, peer ratings, and an instructor evaluation collected during basic training. The criterion was high Airman Performance Rating vs low rating or discharge. Samples were drawn from 15 career fields. Predictive

equations were derived for the full population and for each career-field sample. In all but two career fields prediction was improved by equations based on the career-field samples, but a full-population equation was judged more immediately useful.

41. Flyer, E. S. Prediction of unsuitability among first-term airmen from aptitude indexes, high school reference data, and basic training evaluations (PRL-TDR-63-17). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, June 1963.

Three sets of information were evaluated as predictors of unsatisfactory airman performance as defined by a combination of supervisory ratings and unsuitability discharges: selection and classification variables, basic training performance ratings, and high school reference data. Two 2,000-case samples were identified for which predictor and performance criterion data were available after 2 years of service. Multiple regression analysis applied to the data demonstrated that, within the framework of the current selection and classification process, improved predictions of airman performance are obtainable from educational reference data and behavioral evaluations collected during training. It appears possible to evaluate new airmen during their first month of active duty with a fair amount of accuracy in terms of their potential worth to the Air Force.

42. Flyer, E. S. Unreliable airmen in high-risk jobs: Unsuitability in the munitions and weapons maintenance career field (WADD-TN-60-43). Lackland Air Force Base, TX: Personnel Laboratory, Wright Air Development Division, March 1960.

Lack of adaptability screening in procuring personnel for high-risk positions has resulted in some unreliable personnel being assigned to nuclear weapons duties. In addition, some airmen are maintained in nuclear positions after numerous incidents showing instability or irresponsibility. Techniques are available to screen airmen prior to and during assignment to high-risk positions. While unauthorized nuclear detonation will not be precluded by the most intensive personnel screening, many unreliable airmen will be identified and removed from assignments to high-risk career fields.

43. Flyer, E. S. Working paper on unsuitable airmen: Research investigations by the personnel laboratory during 1961. Lackland Air Force Base, TX: Personnel Laboratory, Wright Air Development Division, April 1961.

This report describes a research program set up by the Air Force Personnel Laboratory for the purpose of developing suitability screening procedures for airmen useful at the recruiting level and during early military training. In this research, large-scale follow-up studies of basic airmen were conducted. As a result, a number of factors were identified that differentiate between those who are successful and unsuccessful in adapting to Air Force life.

Several devices or measures were developed to forecast unsuitability. One device combined in a single score preservice educational level, aptitude, and age information. This single measure identified groups with successful performance rates as high as 85% and as low as 15% for the initial four-year enlistment.

A peer rating device was also developed for use during basic military training that possesses considerable validity in forecasting unsuitability and marginal performance on the job.

44. Frank, B. A., & Erwin, F. W. The prediction of early Army attrition through the use of autobiographical information questionnaires (ARISS TR-78-A11). Washington, D. C.: Richardson, Bellows, Henry, & Company, Inc., July 1978.

Experimental autobiographical questionnaires, including items validated in previous research and new items suggested by that research, were administered to samples of incoming Army enlistees at Forts Dix and Sill. Data on 180-day attrition were collected. Questionnaire results were item analyzed using attrition criterion, and cross-validities were computed for all items. Results substantially confirmed the earlier research outcomes and indicated that autobiographical information could assist in identifying enlisted most likely to experience early attrition.

The previous research is reported in ARI's TR-77-A6, "The Feasibility of the Use of Autobiographical Information as a Predictor of Early Army Attrition."

45. Fruchter, B. Prediction of airman success from responses to items of the Kelley Activity Preference Report (PRL-TDR-62-9). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, June 1962.

Items from a self-report inventory of personal background and activity preferences were selected by various methods and combined to predict successful completion of first-term enlistment. Two samples of airmen (2000 each) were used, each divided into a success group and a non-success group for item analysis and validation purposes. Selection and weighting of valid items was determined on the initial sample; the scoring procedures were cross validated on the second sample. Although optimal item weighting produced higher validity with the initial sample, unit weighting of the most valid items proved as effective in cross validation.

46. Fuchs, E. R., Woods, I. A., & Marpar, B. P. Prediction of job success in eight career ladders (Research Report PRB 997). Washington, D. C.: Personnel Research Branch, TAGO, DA, February 1953.

The Army Classification Battery was being evaluated in several series of studies with the aim of increasing its effectiveness in classifying men. This report describes one of these series in which the evaluating measure of the ACB was effective in predicting on-the-job success. The project was concerned with selecting men for jobs in eight career ladders--Instrument and Fire Control Maintenance, Radar Repair, Fixed Station Radio Repair, Cartography, Counterintelligence, Food Inspection, Preventive Medicine, and Central Office Installation. The selection scores earned by enlisted men on the ten tests of the ACB, the ten Aptitude Areas, and on other combinations of the ACB tests were compared with ratings of job proficiency completed by job supervisors and associates.

The ACB test or combination of tests that best predicted job success was not always the same from job ladder to job ladder. Such evidence of differential selection indicated that the Battery is able to distinguish the ability requirements of different jobs. This finding supported the original intention of the Aptitude Area system to classify personnel on the basis of several special abilities tests rather than depending upon a single, less discriminating, general ability test. For six out of the eight ladders studied, the Aptitude Area currently used to select personnel was one of the best, if not the best, predictor of all the available Aptitude Areas. The results of this series of studies also offered suggestions for profitable reconstitution of some of the Aptitude Areas to increase the differential prediction of job success.

47. Glickman, A. S., & Kipnis, D. Theoretical considerations in the development and use of a non-cognitive battery, Tri-Service Conference on Selection Research, 9-19, Washington, D. C.: ONR, 1960.

The selection of those most able in academic training does not insure the selection of those most able to accomplish the performance requirements of their job. Tests that tap new sources of predictive variance are needed to create the most advantageous school output-fleet input ratio. The problem of prediction lies in selection and training procedures that elevate proficiency level but reduce variation in proficiency, resulting in a situation in which there are few opportunities for observing differences between good and poor workers and making evaluative comparisons. Supervisors, in their search for differentia, have been driven to using non-cognitive factors in making their comparisons. As a result, several non-cognitive tests were developed based upon assumptions concerning characteristics of enlisted men considered by supervisors when judging performance.

If these new tests prove valid, the major problem then becomes one of combining them with the basic Test Battery to ensure efficient selection in terms of both school and duty performance. What is required is a means of weighting tests so that job performance levels are raised while at the same time retaining the advantages in controlling attrition from training schools.

48. Gordon, M. A., & Flyer, E. S. Predicted success of low-aptitude airmen (PRL-TDR-62-14). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, August 1962.

This study examines the performance characteristics of a group of low-aptitude airmen who entered the Air Force during the first 6 months of 1956 and who either completed successfully a 4-year enlistment or were discharged for unsuitability or nonadvancement. It was found that a brief composite of aptitude tests and preservice educational level differentiated the successes from the failures quite well. When it is necessary to recruit from low-aptitude airmen, the additional screening would select those most likely to be of value to the Air Force.

49. Gordon, M. A., & Botenberg, R. A. Prediction of unfavorable discharge by separate educational levels (PRL-TDR-62-5). Lackland Air Force Base, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, April 1962.

Many airmen meet enlistment standards but are nevertheless discharged for unsuitability or failure to advance. A more precise means of identifying men not likely to succeed in the Air Force is needed. This study tested the

hypothesis that different combinations of tests might be needed for men with little schooling than for those at a higher level of education. Multiple regression analyses of the data for two large samples of airmen showed little gain in accuracy of prediction by separate composites for three educational levels. Of the individual predictors of Air Force success, amount of education proved the most valid, further justifying the Air Force in limiting recruitment to high school graduates.

50. Could, R. B., & Sellman, W. S. A biographical inventory as a predictor of test item writer success. Proceedings of the 12th Annual Military Testing Association Conference (1970), 84-95.

The purpose of this paper was to report on a method being developed to identify those noncommissioned officers who will best perform as subject-matter specialists for test item writing duties. A biographical inventory has been developed as a potential selection instrument. The emphasis of the paper is on the method used to develop scoring keys for the inventory and the predictive validity of those keys. The author stated the importance of having qualified subject-matter specialists to develop the Specialty Knowledge Tests and Promotion Fitness Examinations as well as the importance of the tests to the Weighted Airman Promotion System. In an attempt to develop an SMS selection system, a performance rating scale was used to evaluate a biographical inventory, which was developed as a potential selection instrument. The inventory yielded validities in the low .80s and cross validities in the low .50s. The potential value of the inventory as a selection instrument was established and the future use of biographical inventories as prediction instruments should be further explored using the powerful key development system explained in this study. Considering the higher predictive validities obtained in this study as contrasted to past studies, the higher validities may have been obtained not by virtue of an especially effective inventory but rather by the system used to develop the keys. Certainly further studies are indicated.

51. Greenstein, R. B., & Hughes, R. G. The development of discriminators for predicting success in Armor Crew positions (Research Memorandum 77-27). Fort Knox, KY: Army Research Institute for the Behavioral and Social Sciences, December 1977.

The primary focus of this study was on Armor Crewman criterion measures and their relation to potential armor crewman performance predictors. One hundred thirteen armor trainees were used as subjects in order to evaluate the relation between "predictor" variables and AIR performance measures in driving, loading, and firing. Although the results support the relevance of certain predictors for driving and loading criterion performances, it was felt that it would be premature to associate these criterion performances with particular "abilities" (combinations of predictor tests). Instead, the results can be viewed as broadly indicating the existence of empirically identified relations between a class of predictor variables and criterion performance in driving and in loading. Of particular significance in this study was the identification of the statistical independence of driving, loading, and firing performances.

52. Guinn, N., Wilbourn, J. M., & Kantor, J. E. Preliminary development and validation of a screening technique for entry into the security police career field (AFHRL-TR-77-38). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, July 1977.

A sample of 4,502 basic airmen, assigned to the security police career field were administered an experimental battery consisting of biographical, attitudinal, and interest items. Attitudinal scores and criterion data (in/out of service after completion of technical training) were retrieved from the airman record files. Multiple linear regression analyses were accomplished to determine the utility of attitudinal and inventory data in predicting adaptability to the security police career field. The multiple correlations of the final selector composites derived from this study were .46 and .47. Since the small number of enlistees in the sample precluded cross-application of regression weights, it was recommended that further validation be accomplished to determine the reliability and stability of the predictor composites.

53. Guinn, N., Johnson, A. L., & Kantor, J. E. Screening for adaptability to military service (AFHRL-TR-75-30). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, May 1975.

A sample of 15,252 basic airmen were administered the history opinion inventory (HOI) during basic military training. The service careers of these subjects were monitored for two years in order to assess the ability of the HOI to predict the criterion of in/out of service. An a priori adaptation index developed from HOI items correctly identified as high risk 23 percent of those subjects discharged from service during the two-year period, while incorrectly labeling as high risk only 6 percent of those subjects still in service after 2 years. The possibility of increasing the accuracy of prediction by utilizing biographic/demographic data and the operational usefulness of the HOI is discussed.

54. Guinn, N., Kantor, J. E., Magness, P. J., & Leisey, S. A. Screening for entry into the security police career field (AFHRL-TR-77-79). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1977.

A sample of 4,502 airmen assigned to the security police field were administered a test battery consisting of biographical, attitudinal, and interest measures. Using a criterion of in/out of service after a minimum period of 1 year on the job, regression analyses were accomplished to determine the effectiveness of the predictor composites. Efforts were made to decrease the magnitude of the selection composite by eliminating one or more of the experimental test measures or minimizing the overall number of test items. Three selection composites containing different numbers of test items were developed and evaluated for practical utility in identifying individuals most likely to separate from service. The multiple correlations ranged from .46 to .37. Cross-application analyses resulted in multiple correlations of .20 to .19. Recent changes and improvements in this career field were reviewed, and the advisability of implementing a new screening methodology discussed.

55. Gunderson, E. K. E., & Nelson, P. D. Biographical predictors of performance in an extreme environment (Interim Report 65-7). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, 1965.

In this study, the authors assessed relationships between biographical data and performance evaluations for Navy participants in the United States Antarctic Research Program. Prior to deployment to Antarctica, 425 Navy men completed a biographical questionnaire eliciting information concerning military record, interests and hobbies, family and educational background, and vocational experience. After approximately 1 year at an Antarctic scientific station, performance evaluations were obtained from station supervisors and peers. Results from earlier samples (predominantly from large stations) indicated that age, rank, years of naval experience, marital status, worship, delinquency, and amount of reading were significantly related to peer evaluations of adjustment. Results from small-station groups, analyzed in the present study, reveal important differences in the attributes that are correlated with performance criteria.

56. Harding, F. D., & Bottemberg, R. A. Contribution of status factors to relationships between airmen's attitudes and job performance (ASD-TN-61-147). Lackland Air Force Base, TX: Personnel Laboratory, Aero-nautical Systems Command, November 1961.

Previous investigation has shown little relationship between self-report measures of an airman's attitudes (morale) and his rated job proficiency. The data of one such study were reanalyzed by a multiple regression technique to determine whether military status variables (military rank, length of service, kind of duty) affect correlation of attitude measures with proficiency ratings. The addition of such variables to the attitude variables contributed significantly to prediction of supervisors' ratings of proficiency; but the attitude variables did not significantly increase prediction from the status variables alone. The findings show the importance of considering personal and situational factors when evaluating effects of attitude and morale.

57. Hausman, H. J., & Strupp, H. H. Non-technical factors in supervisors ratings of job performance. Personnel Psychology, 1955, 8(1), 201-217.

Supervisory ratings were analyzed to determine whether supervisors could rate aircraft mechanics on several dimensions of on-the-job proficiency, or whether their ratings contained nothing more than global general impressions. Three samples of USAF aircraft mechanics were administered a test of technical competence, while supervisory and co-worker ratings were obtained on them. The rating instruments contained a large number of detailed items, plus one over-all item. Special effort was directed at items that reflected motivational areas, or those "non-technical" variables that appear to modify

job performance so that technical skill does not perfectly predict actual performance.

Several clusters of rating dimensions were obtained from both the supervisory and co-worker rating instruments. These dimensions tended to correlate differentially with measures most likely to reflect technical skill, so that the ratings of technical skill correlated highest with them, while the other dimensions had negligible correlations with test and experience. Rater agreement was at the level usually found for rating instruments, although evidence showed supervisors were better raters than co-workers.

Using a criterion of grouped co-worker overall ratings, it was found that the supervisory rating dimensions were a useful addition to the test of technical competence in predicting overall proficiency. This was interpreted as evidence of the existence of nontechnical variables in the supervisory ratings, since technical competence was covered in the test.

58. Helme, W. H., & White, R. K. Prediction of on-job performance in AAA Gun Crew specialties (PRB Technical Research Note 88). Washington, D. C.: Personnel Research Branch, TAGO, DA, January 1958.

This study was one of a series evaluating the Army Classification Battery (ACB) for effectiveness in predicting success on the job. The objective was to determine the effectiveness of current selectors and a possible alternate for two related jobs in Antiaircraft Artillery--Gun Crewman and AAA Operations and Intelligence Specialist. Scores on the Army Classification Battery tests and on the Aptitude Area test composites of the ACB were compared with ratings of job performance obtained from on-the-job supervisors of 626 enlisted men in MOS 162 and of 38 enlisted men in MOS 163 during the latter half of 1955. The current selectors, Combat A and Combat B, were quite satisfactory for these jobs. These findings indicated that no immediate change of Aptitude Area should be recommended, because any improvement of prediction would require the introduction of new tests in the ACT. Two new measures are already scheduled to be added to the ACB in 1958 to improve prediction of success in combat jobs generally.

59. Helme, W. H., & White, R. K. Prediction of on-job performance in Guided Missile Crew specialties (PRB Technical Research Note 89). Washington, D. C.: Personnel Research Branch, TAGO, DA, February 1958.

This study was one of a series evaluating the Army Classification Battery (ACB) for effectiveness of the current selector and a possible alternate for predicting performance in Guided Missile Crewman specialties--MOS 220, Guided Missile Crewman, and in MOS 225, Surface-to-Air Missile Launching Crewman. Scores on the Army Classification Battery tests and on ACT test

composites were compared with ratings of job performance obtained from on-job supervisors of 231 enlisted men in MOS 220 and of 133 enlisted men in MOS 225 during the latter half of 1955. The EL Aptitude Area currently in use for these specialties was reasonably valid. The predesignated Aptitude Area Combat B was sufficiently valid to be used, if necessary. Validity coefficients of the best composites were generally in the range of .35 to .45 for both jobs--quite high for predicting ratings of job performance--but much lower for predicting performance of NCOs in MOS 225, presumably because experience and leadership requirements are more important than original aptitude for technical duties of NCOs in combat units. No change of selector is to be considered on the basis of these findings, until impending changes in the combat Aptitude Areas have been effected.

60. Helme, W. H., & Boldt, R. F. Prediction of success in selected precision and automotive maintenance jobs (PRB Technical Research Note 98).
Washington, D. C.: Personnel Research Branch, TACO, DA, October 1958.

Validation of current aptitude area selectors and predesignated alternate aptitude areas was carried out for performance in ten enlisted maintenance jobs, four in the Precision Maintenance Occupational Area and six in the Automotive Maintenance Entry Group of the Motor Maintenance Occupational Area. The current selector for the former jobs, Aptitude Area GM, had validity coefficients of .15 to .41. The predesignated alternate areas, Aptitude Area MM for two jobs and Aptitude Area CL for one, had consistently higher validity than did GM, with coefficients ranging from .21 to .45. The current selector for the automotive maintenance jobs had validity coefficients from .15 to .26, a rather low level of prediction but consistently superior to that of the alternate area. In view of these findings and of previous findings from studies of the validity of these selectors for final grades in the prerequisite courses, the following recommendations were made:

That the selector for one MOS, Ammunition Storage Specialist, be changed from GM to CL;

That the study of the feasibility of changing the jobs of Fire Control Instrument Repairman and Turret Artillery Repairman from GM to MM be undertaken;

That research efforts to improve prediction of automotive maintenance jobs be made while use of Aptitude Area MM as selector is continued.

61. Helme, W. H., & White, R. K. Validation of experimental aptitude tests for Air Defense Crewmen (PRB Technical Research Note 90). Washington, D. C.: Personnel Research Branch, TAGO, DA, February 1958.

This study was one of a series evaluating the Army Classification Battery (ACB) for effectiveness in predicting success on the job. The objective of this study was to determine the effectiveness not only of the current selectors, but of six experimental electronic aptitude tests measuring motor coordination, perceptual speed, non-verbal reasoning, and mechanical knowledge, in the effort to improve classification techniques for electronics jobs which are on the increase in importance and number in the Army. Scores on the six experimental tests, on the Army Classification Battery tests, and on ACT test composites were compared with ratings of job performance obtained from on-job supervisors of over 1000 enlisted men as follows: 651 in MOS 162, AAA Gun Crewman; 33 in MOS 163, AAA Operations and Intelligence Specialist; 231 in MOS 220, Guided Missile Crewman; and 133 in MOS 225, Surface-to-Air Missile Launching Crewman. Two experimental composites--Two-Hand Coordination plus Mechanical Knowledge, and Mechanical Knowledge plus Arithmetic Reasoning--had validity comparable to that of the Aptitude Area selectors for these jobs. The results of this study, viewed in comparison with results of combat arms selection studies recently completed, suggested that either of the newly developed Combat Aptitude Area composites was likely to be effective for these jobs. Results also gave some support for the introduction of the Mechanical Knowledge Test into the ACB, possibly as a substitute for the current Shop Mechanics Test.

62. Hickerson, K. A., Hazel, J. T., & Ward, J. H., Jr. A causal analysis of relationships between performance and satisfaction in eight airman specialties (AFHRL-TR-75-57). Lackland Air Force Base, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, October 1965.

Longitudinal relationships, between two measures of both job performance and job satisfaction over a three-year period, were investigated for 1,352 airmen in eight enlisted Air Force occupational specialties. Cross-lagged panel correlation analyses were compared to conclusions based upon an extended multiple linear regression analysis technique. Data presented suggest a causal influence between performance and satisfaction in two of the eight specialties. Other results indicated that the performance-satisfaction relationship is a complex one, dependent upon the models used for investigation, the satisfaction, performance, and moderating variables selected, and the particular job specialty under consideration. The report includes a presentation of the linear regression models employed in the analysis, and a bibliography of performance-satisfaction research.

63. High, W. S., & Mackie, R. B. A factor analytic study of aptitudes, interests, and practical performance skills of Navy Machinery Repairman students (Technical Report VII). Los Angeles, CA: Human Factors Research, Inc., June 1957.

This report describes the results of a factor analytic study to determine the factor content of sixty job sample performance measures and instructor rankings obtained from a sample of 200 Machinery Repairmen students at a Navy Class "A" trade school. The factorial nature of these measures is described in terms of their relationship to fifty-four well-known standard reference tests. Two analytic rotational methods (one oblique and one orthogonal) were used and evaluated in terms of the results obtained with each.

64. Hoiberg, A., & Pugh, W. M. Predicting Navy effectiveness: Expectations, motivation, personality, aptitude, and background variables (NHRC 77-53). San Diego, CA: Naval Health Research Center, November 1977.

The purpose of this study was to identify predictors of performance within seven Navy occupational groups. Life history, expectations, motivation, personality, and aptitude variables were used as predictors of a 2-year effectiveness criterion for 7,923 enlisted Navy men and women. Results of multiple regression analyses showed that the most powerful predictors included: years of schooling, school expulsions and suspensions, the two Comrey Personality Scales of Social Conformity and Orderliness, arrests, age, General Classification Test (aptitude), and Peer Cohesion (expectations). Comparisons across groups indicated that the development of separate equations for each occupation was not supported. Recommendations were made to improve selection procedures and to change several aspects of the organization, suggestions which would be expected to increase rates of effective performance.

65. Johnson, C. D., Burke, L. K., Loeffler, J. C., & Drucker, A. J. Prediction of the combat proficiency of infantrymen (PRB Technical Research Report 1093). Washington, D. C.: Personnel Research Branch, TAGO, DA, July 1955.

More effective procedures are needed for identifying men who will perform successfully on combat jobs and for identifying, before combat, soldiers with fighter potential. The Army's classification system must therefore include better measures of combat potential. Information on the development of Aptitude Areas for classifying personnel into the Combat Arms is presented in this report.

Potentially useful test materials for predicting combat success were tried out in field studies of men in Korean combat and on maneuvers. The most promising results to date have been with self-description measures of the personal traits and attitudes characteristic of the effective combat man. A self-description instrument has been developed which is being integrated with related personal measures and with ability tests in current combat classification research.

66. Johnson, C. D., & Kotula, L. J. Validation of experimental self-description materials for general and differential classification (PRB Technical Research Note 95). Washington, D. C.: Personnel Research Branch, TAGO, DA, August 1958.

Development of self-description measures for enlisted classification was undertaken in the effort to add to the Army Classification Battery more test content that would indicate how well a man will apply himself on the job. Eleven experimental measures were administered to over 1500 cooks, clerks, and mechanics and validated against performance ratings by superiors and peers. Analyses were conducted to determine how well these measures developed for specific job areas predicted job performance in these areas and how well they predicted job performance in general. The measures were found to have usefully high validity coefficients. While the validity of the measures tended to remain high across several different jobs, some promise of differential validity did emerge from further analysis. As a result of the findings of this study, further research in this area will emphasize content with specific orientation toward particular jobs.

67. Judy, C. J. A regression analysis of one set of Airman Proficiency Test scores (WADD-TN-60-139). Lackland Air Force Base, TX: Wright Air Development Division, Air Research and Development Command, June 1960.

One criterion for airman skill upgrading in the Air Force is met by attaining a qualifying score on an applicable Airman Proficiency Test (APT). This note reports an analysis showing the proportion of variance one such test had in common with selected measures of training, experience, education, aptitude, supervisory opinion, and airman attitudes for a sample of 384 aircraft mechanics tested in 1956 and 1957. Each of these categories of information, excepting airman attitudes, could be used to predict the APT criterion at some level of effectiveness; but only the training variables and the aptitude variables added significantly to the prediction attainable by using all other available information. Other research was cited in which various APT correlates were reported. Results showed the utility of APT scores in defining one important aspect of airman proficiency.

68. Kantor, J. E., & Guinn, N. Comparison of performance and career progression of high school graduates and non-graduates in the Air Force (AFHRL-TR-75-73). Lackland Air Force Base, TX: Air Force Human Resources Laboratory, Personnel Research Division, December 1975.

The performance and career progression of a sample of 20,705 air-men were monitored throughout their initial tour of service. For comparative purposes, this sample was divided into high school graduate and non-graduate groups and further subdivided by Armed Forces Qualification Test (AFQT) mental categories. Points of comparison included: disposition from basic military and technical training, attainment of skill levels, number of disciplinary actions and unsuitability discharges, and reenlistment decision. On almost all measures, high school graduates constituted a significantly more successful military group than did the non-graduates, and among the non-graduates, in terms of mental category subgroups, there were almost no differences in performance. In addition, the effects of varying enlistment requirements on this sample are presented, and attention was directed toward determining which non-graduates might be better risks than others for military service.

69. Kipnis, D., & Glickman, A. S. Development of a non-cognitive battery: Prediction of performance aboard nuclear powered submarines (Technical Bulletin 61-5). Washington, D. C.: Bureau of Naval Personnel, February 1961.

In 1958, several of the non-cognitive tests (Hand Skills, Error Finding, Color Naming) were administered to two incoming classes of enlisted men at the Naval Nuclear Power (NP) School, New London, Connecticut. Two years later, evaluations of performance aboard NP submarines were obtained for 117 men from these classes. Nine areas of performance were evaluated. This report gives the validities of the non-cognitive tests against these evaluations.

For each of the performance areas, the 117 men were divided into those categorized as Below Average (lower third) and those Average or better. Biserial correlations were computed between the dichotomized criteria and the three non-cognitive tests, the Basic Test Battery (BTB), and eighth-week academic grade at Basic NP School. The three non-cognitive tests were significantly related to performance evaluations. The Error Finding Test had validities ranging from .09 to .35, with a median validity of .25. The Hand Skills Test most clearly predicted evaluations of technical competence, and Part I of the Color Naming Test predicted evaluations of both technical and non-technical performance. The best two-test combination was the Hand Skills plus Error Finding.

Out of 27 possible correlations between the three tests of the BTB and the nine performance evaluation areas, only one significant prediction was obtained between GCT and Ability to Maintain Equipment. On the other hand, Basic NP School Grade was highly predictive of all components of performance, except Military Appearance. Useful predictions of duty performance aboard NP submarines were obtained with the non-cognitive tests. It would appear that a test battery made up of the Hand Skills Test and the Error Finding Test would give a good prediction of duty performance. Further research with revised versions of these tests is being carried out with additional groups of NP applicants.

70. Kipnis, D., & Glickman, A. S. The development of a non-cognitive battery: Prediction of radioman performance (Technical Bulletin 59-14). Washington, D. C.: Naval Personnel Research Field Activity, June 1959.

This is the third in a series of reports on the development of tests of temperament and personality for predicting the performance of enlisted personnel. In this study a number of experimental tests of temperament and personality were administered to entering students at Class A Radioman School. Final grades and marks in code-retrieving were obtained from school records. Evaluations of duty performance were obtained from each Radioman's most immediate supervisor about one year after testing.

It was found that some of the experimental tests were reasonably effective in predicting both the school grades and the level of performance of duty of the enlisted man. Because of the promise indicated in this study, the research with tests of temperament and personality is being continued and expanded to include different ratings.

71. Kipnis, D., & Glickman, A. S. The development of a non-cognitive battery to predict enlisted performance (Technical Bulletin 58-9). Washington, D.C.: Naval Personnel Research Field Activity, August 1958.

The purpose of this report was to present the results to date of a program to develop non-cognitive predictors of enlisted performance. Six experimental tests were developed based upon assumptions concerning characteristics of enlisted men considered by supervisors when judging performance. The tests were administered to a sample of 125 third class Aviation Machinist Mates (AD3s). Evaluations of performance were obtained by unofficial evaluations of performance made by each AD3s' leading petty officer. For purposes of analysis, two sub-samples (A and B) were formed from the total number of AD3s. Product-moment intercorrelations of predictors, and biserial correlations of each predictor with the criterion, were computed.

Three of the six tests (Color Naming, Hand Skills, and Risk Scale) showed significant correlations with supervisory evaluations of performance at the .05 confidence level. Two other test validities (Error Finding and Sports Scale) came close to this. Individual test validities ranged from -.22 to .32 for Sample A, and from -.32 to .49 for Sample B. Multiple correlations of .500 and .667 were obtained using all five tests. Scores from the Basic Test Battery showed no relationship with the criterion.

72. Kipnis, D. A noncognitive correlate of performance among lower aptitude men. Journal of Applied Psychology, 1962, 46(1), 76-80.

The Hand Skills Test, a device that measures "persistence beyond minimum standards on tiring tasks," was used to predict school grades and job performance evaluations for higher and lower aptitude Navy personnel. Three enlisted samples and one officer candidate sample were employed. Within each sample men were divided into higher and lower aptitude groups at the median of their aptitude test scores. Principal findings were: (a) the Hand Skills Test significantly predicted school grades of the two lower aptitude enlisted samples (grades were not available for third enlisted samples) but did not predict for higher aptitude enlisted men or for officer candidates and (b) the Hand Skills Test significantly predicted job performance evaluations among lower aptitude men in all four samples, but again validities were not significantly different from zero among the four higher aptitude samples.

73. Kipnis, D. Prediction of job performance. Journal of Applied Psychology, 1962, 46(1), 50-56.

A battery of noncognitive tests was developed to improve prediction of Navy enlisted men's performance evaluations. Reported are the results of one concurrent validity study and two follow-up studies with intervals of 14 and 30 months between testing and performance evaluations. Ss were 125 aviation machinist mates, 128 radiomen, and 117 nuclear power personnel. The study revealed: (a) the experimental tests were independent of the Navy's Basic Test Battery, with the exception of the speeded clerical coding test; (b) the tests were most efficient in identifying men categorized as Below Average in performance; (c) tests attempting to measure persistence beyond minimum standards, decisiveness, and lack of insolence yielded significant prediction of performance. Composite validities about .40 were obtained in the two follow-up studies.

74. Kipnis, D. Prediction of performance among lower aptitude men (Technical Bulletin 61-10). Washington, D. C.: Bureau of Naval Personnel, July 1961.

The purpose of the study was to investigate whether the Hand Skills Test, constructed as an attempt to measure "persistence beyond minimum standards on tiring tasks," was an equally valid predictor of school grades and job performance evaluations among higher and lower aptitude men.

Three samples, consisting of from 122 to 135 enlisted radiomen, from 117 to 240 enlisted nuclear power men, and 108 officer candidates, were employed. GCT scores were used as the measure of aptitude for enlisted men and OQT scores for officers. Within each sample, men were divided at the median of their aptitude scores into High and Low Aptitude Groups. The principal findings were:

1. Among radiomen and nuclear power men, the Hand Skills Test significantly predicted the school grades of low aptitude men with phi coefficients of .23 and .29 respectively. Among high aptitude radiomen and nuclear power men, test validities were not significantly different from zero. The Hand Skills Test did not predict school grades among either high or low aptitude officer candidates.
2. The Hand Skills Test significantly predicted job performance evaluations among low aptitude men in all three samples. Phi's ranged from .26 to .47. On the other hand, none of the validity coefficients were significantly different from zero among high aptitude men in all three samples.
3. The findings suggested that as group aptitude level decreased, the validity of the Hand Skills Test increased.

75. Kipnis, D. The relationship between persistence, insolence, and performance, as a function of general ability. Educational and Psychological Measurement, 1965, 25(1), 95-110.

Two hypotheses were tested in this study: (1) Persistence will be positively related to school and job performance among lower ability men and not related to performance among higher ability men; (2) Insolence will be negatively related to performance among higher ability men and not related to performance among lower ability men.

Over 1700 Navy recruits were tested, using the Hand Skills test to measure persistence, the Insolence Scale, to measure passive-aggressivity, and the General Classification Test to measure general intelligence. Performance criteria used were trade school grades and job performance evaluations by superiors.

Results did not generally support the first hypothesis. On the other hand, job performance analysis did provide some support for the second hypothesis. A possible explanation for hypothesis two results is that they represent an interaction between personality and task difficulty. It may be that higher ability men who are also high in insolence become rapidly bored with their work and express their boredom by directing hostility toward authority. If this is the case, one might consider placing this type of individual in work that is somewhat challenging and difficult for him. Assigning these individuals to easier work may lead to a failure to realize their full potential because they become "cocky" and possibly indifferent to the work assigned.

76. Klieger, W. A., Dubuisson, A.V., & Sargent, B. B., III. Correlates of disciplinary record in a wide-range sample (Technical Research Note 125). Washington, D. C.: Army Personnel Research Office, August 1962.

Means of identifying potentially delinquent soldiers are being developed on current Army input-restricted, in effect, to the upper three AFQT mental categories. Would such means be effective also in a mobilization input that included men in AFQT categories IV and V?

Operational test scores and data on type of discharge and court-martial conviction were obtained on a sample of 875 enlisted men who entered the Army in 1952-53 when AFQT IV and V men were being accepted. High and low AFQT categories were compared with respect to disciplinary action and predictors were evaluated in the broad-based sample.

AFQT IV and V categories showed significantly greater proportions of men incurring disciplinary action than did AFQT category III and above. Years of education, the verbal test of the Army Classification Battery, and AFQT were consistently related to the disciplinary criterion in a sample in which all mental categories were represented. Pre-service criminal record was also related to disciplinary action. In view of criterion differences established between a broad-based (mobilization) sample and a restricted (current input) sample, age at entry or score on a specially developed predictor could be considered as additional qualifying factors for use with applicants or registrants in AFQT IV and V categories.

77. Klieger, W. A., de Jung J. E., & Dubuisson, A. U. Peer ratings as predictors of disciplinary problems (USAPRO Technical Research Note 124). Washington, D. C.: U. S. Army Personnel Research Office, July 1962.

Satisfactory means are needed to identify incoming soldiers who meet current induction or enlistment standards but whose Army performance is likely

to prove unacceptable. Peer and cadre ratings during basic training were among available measures that needed evaluation as possible predictors.

In this study, discharge, court-martial, and promotion records covering two years of service (three years in the case of three-year enlistees) were obtained for 1,571 enlisted men entering the Army in 1955. Ratings obtained during basic training as well as test and background data were evaluated as predictors of behavior warranting disciplinary action.

Findings were that ratings of combat potential made as early as the 5th week of basic combat training showed substantial validity in predicting acceptability. Since the ratings showed higher validity than any of the other experimental predictors of disciplinary problems, further exploration of their utility for this purpose is desirable.

78. Kliager, W. A., Dubuisson, A. U., & de Jung, J. E. Prediction of unacceptable performance in the Army (HFRB Technical Research Note, No. 113). Washington, D. C.: Human Factors Research Branch, TAG R&D Command, DA, June 1961.

In response to a requirement for early identification of those enlisted personnel likely to become disciplinary problems in the Army, a number of personnel measures were evaluated as possible predictors of military unacceptability. Proficiency and performance test scores, AFQT scores, Average Basic Training Ratings, and various indices based upon background information were obtained on a group of 1,780 first-term enlistees completing basic combat training at Fort Leonard Wood during 1953-54. Indices of unacceptability were based on type of discharge and courts-martial record. A composite of AFQT score and age at entry provided highest prediction of the acceptability criterion (multiple $R = .41$). Of the other measures, only the Ratings and the stakes test (performance measure) added to the predictive-ness of the composite. Findings are not final ones, as additional data on prediction of unacceptable performance are still being analyzed.

79. Knapp, R. R. Prediction of recidivism from peer ratings, self ratings, and personality inventory factors (Report 161). San Diego, CA: Personnel Research Field Activity, February 1961.

Personality inventory and sociometric data had been collected during 1955-1956 from a sample of disciplinary offenders at the U.S. Naval Retraining Command, Camp Elliott, California. This report considers the validity of these measures for predicting a criterion of restoration success during a 6-month follow-up period.

Scores from the Gordon Personal Profile and the Gordon Personal Inventory were validated against the actual probation record obtained from a six-month follow-up on the basic sample of 162 to whom these personality measures had been administered. Mean inventory scale scores for the sample were also compared with scores for an unselected sample of basic airmen. For an augmented sample of 412, self-estimates and peer estimates of anticipated probationary success were validated against this probationary success criterion.

For the basic sample to which the inventories had been administered a significant validity was obtained against the personality trait Original Thinking. In the augmented sample the correlations against self and peer estimates also reached significance. Comparison of mean inventory scale scores for the prisoners with those obtained from an unselected sample of basic airmen showed the prisoners to be significantly lower on Ascendancy, Responsibility, Emotional Stability, Cautiousness, and Personal Relations.

Present results indicated that peer and self-estimates of probable restoration success and one personality measure were significantly related to the 6-month follow-up criterion. However, the small magnitude of the correlations led to the conclusion that these measures, as presently studied, would not have operational use in evaluating prisoners. Future studies might profitably investigate the predictive validity of background information believed to be associated with occurrence of disciplinary offenses in the military. In addition, it would be of basic interest to broaden the range of personality variables investigated.

80. Knapp, R. R. The relationship between certain personality measures and delinquency rate in a Navy sample (Technical Bulletin 61-9). Washington, D. C.: Bureau of Naval Personnel, August 1961.

Habitual delinquency is a problem of major concern to operating Navy personnel. Identification of personality characteristics associated with delinquency rate constitutes a first step in the development of measures for use in the early screening out of the habitual or frequent offender. This study was conducted to determine whether personality scales measuring social maturity and conformity were related to delinquency rate in a group of brig confinees. Significant relationships were found between these scales and delinquency rate. Present findings suggest the advisability of investigating the predictive efficiency of measures of these personality characteristics in other military settings.

81. Lau, A. W., & Abrahams, N. M. The Navy Vocational Interest Inventory as a predictor of job performance (Research Report SRR 70-28). San Diego, CA: Naval Personnel and Training Research Laboratory, April 1970.

Little success has resulted from earlier attempts to develop psychological instruments to predict the job performance of enlisted personnel. Given a valid instrument, men could be assigned to those jobs for which their performance will be at a maximum. The purpose of this research is to evaluate the validity of occupational Navy Vocational Interest Inventory (NVII) scales as a predictor of enlisted job performance.

The NVII was administered experimentally to incoming students at seven Class "A" schools varying widely in curriculum. Scores on Basic Test Battery (BTB) subtests and Final School Grades (FSG) were also available as predictors of enlisted performance. Performance measures consisted of scores earned on the Report of Enlisted Performance Evaluation. Multiple correlations were computed between various combinations of predictors, using performance scores as the criterion.

NVII scales were found to be moderately related to performance scores in three of the seven ratings. In general, little or no relationship was found between interest scores and performance nor did the NVII scales supplement BTB scores when the two were combined. Final Class "A" school grades appeared to be more effective in predicting performance than either BTB or NVII scores. Possible reasons for the low NVII validities observed are: (a) job performance criteria not sufficiently differentiating and relevant, and (b) samples too small for development of empirical performance prediction keys.

82. Lecznar, W. B., & Davydiuk, B. F. Airman Classification Test Batteries: A summary (WADD-TN-60-135). Lackland Air Force Base, TX: Wright Air Development Division, Air Research and Development Command, May 1960.

Assignment to training and jobs has been effectively accomplished by the Air Force through the use of test batteries. Two basic testing instruments have been used: the Airman Classification Battery and the Airman Qualifying Examination. These two tests have been revised periodically to counteract item obsolescence incurred by technology changes, to protect test security, and to use new test theory. Revisions in test content, format, and administration also have been prompted by validation studies. This report compiles a review of each form of these tests, together with development information, and citation of published reports.

83. Lockman, R. F. Enlisted selection strategies (CNS 1039). Arlington, VA: Center for Naval Analyses, September 1974.

The efficiency and fairness of procedures used to select enlisted men for the Navy and for schools, jobs, and advancement were examined. The literature on selection-testing, training, and performance evaluation was reviewed. Ways of increasing personal performance and opportunity are suggested.

84. Mackie, R. R. Factors influencing the use of practical performance tests in the U. S. Navy. Proceedings of the 9th Annual Conference of the Military Testing Association (1967), 32-40.

This report describes a survey undertaken in 1962 to determine the degree to which practical performance tests were being used in the Navy, who was using them, and what attitudes existed toward their use. Results showed that, with the exception of the Radioman ratings, the operating forces did not make extensive use of this type of test (or any other kind of formal test) but depended almost exclusively on supervisory judgments as a means of evaluating performance.

The reasons given for not using performance tests were generally that (1) they're not practicable, (2) none are available, or (3) supervisory judgments are better. Serious concern is expressed that, due to a significant underestimation by operational commanders of the degree of training required to effectively operate and maintain today's complex military systems, the

85. Mackie, R. R., Ridhalgh, R. R., & Shultz, T. E. New criteria for the selection and evaluation of sonar technicians (NPRDC TR 81-13). San Diego: Navy Personnel Research and Development Center, July 1981. (AD-B059 973L)

In the interest of improving selection and evaluation procedures for operator personnel of current and future sonar systems, a number of standardized and experimental selection tests were administered to a sample of students undergoing sonar operator training at the ASW Training Center, Pacific. The predictor tests were later validated against typical academic (written test) criteria as well as against measures of operational performance including target detection, report timeliness, target classification, and target tracking and localization. It was shown that presently used selection tests are totally inadequate as predictors of operational performance though they do predict academic performance. Use of a number of the experimental predictor tests would substantially improve the selection process as measured by either academic or operational criteria.

86. Mackie, R. R., Wilson, C. L., & Buckner, D. N. Research on the development of shipboard performance measures: Interrelationships between aptitude test scores, performance in submarine school, and subsequent performance in submarines as determined by ratings and tests (Technical Report V). Los Angeles, CA: Management and Marketing Research Corporation, October 1954.

This report describes research that was conducted to determine the relationships among scores on a variety of aptitude tests, standing in Basic Enlisted Submarine School, New London, CT, and subsequent performance aboard submarines as measured by ratings, written tests, and job sample tests. The interrelationships of the several shipboard performance measures are described and the results of a factor analysis of the intercorrelations of aptitude test scores and Submarine School criteria are presented.

87. Mackie, R. R., & High, W. S. Research on the development of shipboard performance measures: Supervisory ratings and practical performance tests as complementary criteria of shipboard performance (Technical Report IX). Los Angeles, CA: Human Factors Research, Inc., June 1959.

This report describes a shipboard follow-up study of the performance of Navy Machinery Repairmen whose aptitudes, skills, interests and achievements had been thoroughly studied two years earlier while they were in Class "A" MR training. Shipboard performance was assessed by administering a practical performance test requiring skill in the use of machinery repair equipment, and by securing ratings by supervising petty officers of each person's ability to perform the various aspects of the MR's shipboard job.

The results strongly suggested that performance tests and supervisory ratings were best regarded as complementary criteria of shipboard performance. While these two measures did not correlate with each other, both correlated significantly with many logical predictors, including aptitude and interest measures, practical work during training, and predictions of success by Class "A" school instructors.

When the two shipboard measures were combined to form a simple composite criterion, it was estimated that over 50 percent of the true variance was accounted for by scores made two years earlier on a combination of scores made on: (1) mechanical knowledge tests; (2) training projects involving the use of lathe and milling machines; and (3) predictions by school instructors as to eventual suitability as an MR.

88. Maier, M. H., & Fuchs, E. F. Effectiveness of selection and classification testing (ARISS RR 1179). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences, September 1973.

The various testing programs in the Army's enlisted personnel system are described, and the relationships between testing program, training content and method, and utilization on the job are probed. A brief explanation is given of the methodology by which the effectiveness--that is, the validity--of the tests is established. Analysis of measures of performance in job training programs and ratings of performance on the job reveals that training performance is more satisfactory than job ratings for evaluating the effectiveness of selection and classification tests. How well tests predict performance in job training programs and the relationship between test scores and other indexes of success are examined separately for Negroes and whites.

Selection and classification tests through twenty years of research and experience have demonstrated their effectiveness in identifying potential failures in Army training programs and for getting men into jobs where their potential is best utilized and they can best serve the Army. Aptitude test scores are useful indicators of the level of proficiency and grade a man can attain and of the time required to bring a trainee to a minimum level of performance. The tests are related to rate of promotion in the Army and to civilian earnings after separation from service. Much the same order of relationship holds for Negroes and whites.

The present report analyzes the general criticism of tests, and of military tests in particular, that has arisen in recent years. The analysis supports the usefulness of tests in the Army's personnel systems.

89. McFarlane, T., Kantor, J. E., & Guinn, N. Correlates of successful on-the-job performance in the Security Police (Air Force Specialty Code 81XXX) Career Field (AFHRL-TR-79-16). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, June 1979.

A Security Test Battery, tapping pre-training biographic/demographic factors and post-training job experience factors, was administered in the field to 3,175 Security Police (81XXX) personnel. Job performance ratings were simultaneously collected on these personnel from their first-line supervisors. Using multiple linear regression analyses, it was found that 24 pre-training factors were significantly related to job performance. It was possible to categorize these specific items into four major areas: age, attitudes toward parents and former teachers, family's socio-economic status, and aspects of the individual's personal lifestyle. From the post-training job experience factors, 13 significant correlates of job performance were found which could also be grouped into four attitudinal areas: toward supervisors, the Air Force in general, environmental factors, and co-workers. Cross-application of these results indicated reasonable generalizability. The potential effects of manipulating these variables through selection, classification, and management are discussed.

90. McGrath, J. J. Human factor problems in anti-submarine warfare: Cross-validation of some correlates of vigilance performance (Supplementary Note to Technical Report 4). Los Angeles, CA: Human Factors Research, Inc., February 1961.

In an exploratory study of the correlates of vigilance performance a number of significant correlations were found between psychological test scores and measures of vigilance performance. In subsequent studies of vigilance, cross-validation data were obtained and several additional tests were administered. The results showed that none of the 35 test variables studied consistently predicted performance on auditory and visual vigilance tasks. This negative finding was considered to be a reflection of the task-specificity of individual differences in vigilance performance and made questionable the possibility of selecting through the use of traditional psychological selection techniques the more vigilant performers for practical vigilance tasks.

91. McGrath, J. J., Harabedian, A., & Buckner, D. N. Human factor problems in anti-submarine warfare: An exploratory study of the correlates of vigilance performance (Technical Report 4). Los Angeles, CA: Human Factors Research, Inc., February 1960.

The purpose of this study was to investigate the relationship between a large number of behavioral measures (psychological tests, threshold measures, and subjective reports) and criteria of performance on vigilance tasks. The effort was directed toward ascertaining the types of behavioral measures, rather than the specific measurement instruments, that would be promising predictors of vigilance performance.

Major findings were that: (1) None of the psychological tests used in the study were valid enough to be useful by themselves in personnel selection; (2) Tests measuring clerical abilities appeared to be promising predictors of the amount of decrement in detection performance suffered by individuals during watch, but did not appear to predict the overall performance levels; (3) Performance on an auditory vigilance task was more predictable from psychological test scores than performance on a visual vigilance task; (4) There was a significant correlation between brightness discrimination threshold and performance on a visual vigilance task; (5) Subjects detected fewer signals when they reported feelings of tiredness; (6) Qualitative differences in vigilance performance (sleeping vs not sleeping on watch) were more predictable from psychological test scores than quantitative differences in vigilance performance (percentage of signals detected); and (7) The percentage of signals detected on watch was positively related to the amount of sleep the subject obtained the night before watchstanding.

92. Meranda, P. F. Navy petty officer promotion examinations as predictors of on-the-job performance. Educational and Psychological Measurement, 1959, 19(4), 657-661.

A study was made of 40 Navy-wide petty officer examinations for advancement in rating against a criterion of on-the-job performance in the form of a 10-discrete category rating scale. Samples ranged from 28 to 245 and were distributed among three petty officer pay grade levels. Median validity coefficients were respectively .49, .21, and .25 for examinations of petty officers, 1st, 2nd and 3rd class. Twenty-four of the forty validity coefficients ranged from .20 to .70. On the basis of these findings it appears that the Navy-wide examinations for petty officers have sufficient validity to predict on-the-job performance of candidates for promotion to the next higher pay grade.

93. Mullins, C. J., & Winn, W. R. Criterion development for job performance evaluation: Proceedings from Symposium 23 and 24 June 1977 (AFHLR-TR-78-85). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, February 1979.

This report consists of the proceedings from a symposium conducted in San Antonio, Texas. The purpose was to bring together several researchers who have been recently concerned with various aspects of criterion research to exchange ideas over a 2-day period, and to provide discussion and critique of the directions their respective research efforts are taking. More formal presentations of work and ideas connected with criterion research by military scientists comprised the central part of the 2-day period. It was preceded by more informal material in the way of introductory remarks, and it was followed by summary material provided by a panel of five eminent researchers from the civilian community who were invited to serve as expert consultants and to give their views concerning the work.

94. Plag, J. A., Goffman, J. M., & Phelan, J. D. The adaptation of naval enlistees scoring in Mental Group IV on the Armed Forces Qualification Test (NMMRU Report No. 68-23). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, 1967.

This report has presented findings from a study designed to evaluate differences in the adaptations of "average" and mentally marginal sailors during 4 years of military service. Sailors with AFQT scores of 50 are significantly superior to Category IV enlistees on military performance measures in which cognitive abilities play an essential role. While Mental Group IV sailors have appreciably lower rates of overall naval effectiveness, they do not differ significantly from average enlistees with respect to disciplinary and illness rates.

Four pre-enlistment characteristics were found to be valid for predicting 4-year naval effectiveness among Category IV personnel. These four variables were years of schooling completed, number of school expulsions, AFQT score, and number of arrests. An actuarial table, showing the probability of naval effectiveness as a function of different combinations of these four predictors, was constructed as a guide for the use of recruiting officers in making decisions concerning the enlistment of mentally marginal applicants.

95. Plag, J. A., & Hardacre, L. E. Age, years of schooling, and intelligence as predictors of military effectiveness for naval enlistees (NMNRU 65-19). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, July 1965.

The validity of age, education, and GCT score in the prediction of four criteria of 2-year military effectiveness were examined for a group of 952 enlistees who entered naval service in 1960. Subjects were graduated from training without being subjected to routine psychiatric screening procedures. Thus, the findings are applicable as a guide for clinicians at training commands who regularly make decisions concerning the efficacy of service retention for recruits who experience adjustmental difficulties in training.

The four criteria of effectiveness were pay grade level, division officer ratings of adjustment, semi-annual marks, and record of disciplinary or commendatory action. Data for the 952 subjects comprising the validation sample were analyzed by multiple correlation procedures. Regression equations were derived for each criterion and cross-validated on another subject group of comparable size. The relations of each of the three predictors with the four criteria were found to be statistically significant and consistent from criterion to criterion. When combined, each of the independent variables contributed uniquely to the multiple correlations, but these were generally of small magnitude, ranging from .26 to .45 for the cross-validation sample.

Charts were constructed to facilitate the determination of predicted criterion scores from specific combinations of the age, education, and GCT score variables. In addition, the ability of predicted scores to differentiate criterion subgroups and the odds of enlistees with specific predicted scores falling into criterion subgroups were represented graphically.

96. Plag, J. A., Goffman, J. M., & Phelan, J. D. Predicting the effectiveness of new mental standards enlistees in the U.S. Marine Corps (NMNRU 71-42). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, 1970.

This study compares the performance and adjustment of "new mental standards" Marines with enlistees of higher mental ability. About four out of ten new standards Marines fail to complete a 2-year tour successfully, while only one out of four high ability Marines fail to do so. Thirteen of 34 pre-enlistment characteristics, 4 of 12 early training variables, and 5 of 17 later training

variables have significant validities for predicting effectiveness during a first tour of duty for low ability Marines. The best of these predictors have been combined into tables of odds for ready estimation of the chances that a recruit will successfully complete a two-year tour. Use of these tables of odds at recruiting stations could help in the selection for enlistment of Marine applicants most likely to serve effectively.

97. Plag, J. A. Predicting the military effectiveness of enlistees in the U.S. Navy (69-23). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, 1969.

At each of four stages during the first enlistments of Navy personnel, regression equations were derived for predicting military effectiveness. While the composite validities were not sizable, they accounted for a significant percentage of the criterion variance, and were impressive when it is considered that they represented predictions over a 2- to 4-year period. It is of interest to note that the composite predictions of effectiveness made at the termination of recruit training (Stage C) were not a great deal more accurate than those made prior to enlistment (Stage A). In other words, the pre-enlistment adaptation of applicants, as reflected by school adjustment and cognitive ability, accounted for almost as much criterion variance as that which was predictable from knowing enlistees' recruit training performance.

98. Plag, J. A., & Goffman, J. M. The prediction of four-year military effectiveness from characteristics (66-8). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, August 1966.

The purpose of this study was to examine the relation of background characteristics of naval recruits to a 4-year criterion of military effectiveness. Sailors classified as rendering effective performances were those completing their periods of active obligated service and being recommended for reenlistment. Linear multiple regression procedures were employed for the purpose of deriving an equation in which statistically significant predictors would receive optimal weights and yield a predicted score indicative of a subject's probability of naval effectiveness. Such probability estimates, it was reasoned, could be of value to psychiatrists who regularly make decisions to retain or discharge recruits from service.

For the experimental samples, totaling 3,630 sailors, it was found that approximately 73 percent rendered effective service. A combination of five recruit characteristics was found to give the best prediction of the 4-year criterion. These were level of schooling, family stability, number of expulsions from school, Arithmetic Test score, and Mechanical Test score. Data analyses indicated the derived composite multiple prediction of effectiveness to be far more valid than clinicians' judgments at the time of the initial recruit training screening interview. Suggested uses of computed effectiveness probabilities in the Navy's preventive psychiatry program at recruit training commands were discussed.

99. Flag, J. A. Pre-enlistment variables related to the performance and adjustment of Navy recruits. Journal of Clinical Psychology, 1962, 18, 168-171.

For the purpose of investigating the predictive validity of one type of information available to the clinician at the time of the psychiatric screening examination, an experimental inventory, composed of 195 questions, selected as measures of 11 areas of psychological development and pre-service performance, was administered to 20,000 Navy recruits entering training. Through an analysis of a sample of 6,195 cases, contained in a validation and a cross-validation group for differentiating within four criteria of successful recruit performance, seven psychological areas were delineated as containing the greatest number of significant variables. Highly valid predictors were itemized for use as possible standards at recruiting stations for the rejection of applicants with minimal adjustment and performance potential.

100. Flag, J. A., & Goffman, J. M. The utilization of predicted military effectiveness scores for selecting naval enlistees (69-6). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, December 1968.

Based upon the findings of this study, the following conclusions seemed to be warranted: (1) Many more applicants were qualified for enlistment into the Navy each month than could actually be accepted; (2) the selection ratio was more favorable for personnel qualified in AFQT Mental Groups I-III than it was for personnel qualified in Mental Group IV; (3) on the basis of differences in predicted effectiveness scores, the quality of enlistees presently entering the Navy was considerably higher than the quality of enlistees who entered service in 1960; (4) there was no difference in the mean predicted effectiveness scores of applicants currently qualified for enlistment and those actually enlisted. This finding suggested that, because of quota limitations, delays in enlisting applicants did not result in a loss to the Navy of more personnel of high quality than of low quality; (5) although the reliability of predicted effectiveness scores was less than optimal, it was considered to be within the range of acceptability; (6) sufficient variability existed in the predicted effectiveness score of prospective enlistees to warrant the use of these scores for distinguishing between applicants who should be enlisted and those who should not. It was estimated that as many as 4,500 of the non-effective sailors who are currently entering the Navy each year could be eliminated from service if predicted effectiveness scores were used for selecting personnel.

101. Flagg, J. A., & Hardacre, L. E. The validity of age, education, and GCT score as predictors of two-year attrition among naval enlistees (64-15). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, June 1964.

The validity of the predictor variables of age, education, and GCT score and the criterion of 2-year attrition were examined for a group of naval enlistees who entered service in 1960 and graduated from recruit training without being subjected to the process of psychiatric screening. GCT score and level of educational achievement were found to be negatively related to attrition, with the GCT relationship being nearly linear and the education relationship approximating the cotangent function. Age, on the other hand, showed a slight, but distinctly hyperbolic relationship with the criterion, the lowest discharge rates occurring among 18-year-old enlistees. As a result of interaction effects between predictors, it was found that younger enlistees who are high school graduates and possess high GCT scores had the lowest rates of discharge of any group, while highest rates of discharge occurred for enlistees who were also 17 years of age, but who had little schooling and possessed low GCT scores. Probability tables, which can be used for predicting retention from a combination of educational level and GCT score, were constructed separately for each of three age categories.

102. Robertson, D. W., Ward, S. W., & Royle, M. H. Evaluation and prediction of Navy career counselor effectiveness (NPRDC TR 77-35). San Diego, CA: Navy Personnel Research and Development Center, June 1977. (AD-A042 032)

The Navy's Career Counseling Program assigns senior petty officers knowledgeable in the Navy's training and career programs to assist enlisted personnel in taking advantage of relevant career opportunities. Selection procedures were developed to identify senior petty officers who would be most concerned and effective in providing career guidance service. Criterion data were acquired directly from the counselees who evaluated such counselor behaviors as pleasantness, thoroughness, and interest in the counselee's concerns. Noncognitive predictor instruments administered to counselors included the Guilford Tests of Social Intelligence (GTSI), Comrey Personality Scales (CPS), Strong Vocational Interest Blank (SVIB), Dole Ideal Counselor Adjective Check List (ICAC), and a locally developed Biographical and Attitudinal Inventory (BAI). Scoring keys to predict counselor effectiveness were empirically constructed, and standard keys were validated for the GTSI, the CPS, and a cognitive test, the Navy Basic Test Battery (BTB).

Counselees evaluated counselors favorably on pleasantness, concern, and awareness. Younger counselees evaluated counselors in the 32-34 year age range highest, and low aptitude counselees evaluated counselors' helpfulness more highly than did high aptitude counselees. Neither the counselor's seniority level nor BTB scores were related to counselees' evaluations. In

cross-validation of the noncognitive predictors, validities for standard keys ranged from near zero to the low 20s, while validities for the empirically constructed keys ranged from near zero to the 40s. For selection ratios ranging from 30 to 70 percent, use of the keys would yield proportionate improvement of from 8 to 26 percent.

Use of the empirically constructed key for the CPS was recommended for Navy Counselor selection. Further validation of the BAI/ICAC composite key was recommended, as was validation of all three keys for use with other jobs involving counseling activities.

103. Sands, W. A. Development of a revised Odds for Effectiveness (OFE) table for screening male applicants for Navy enlistment (NPRDC TN 76-5). San Diego, CA: Navy Personnel Research and Development Center, April 1976.

The original Odds for Effectiveness (OFE-1) table was designed to estimate the probability that a man would render effective naval service as a function of: (1) aptitude test score, (2) number of years of school completed, (3) number of expulsions/suspensions from school, and (4) number of arrests. However, after the OFE-1 table was implemented in the beginning of 1973, Navy recruiters experienced increasing difficulty in obtaining arrest information. The purpose of this investigation was the development of a revised Odds for Effectiveness (OFE-2) table that would not require arrest information for enlisted applicants.

A sample of persons ($N = 3,649$) entering the Navy in 1960-61 was divided into a development sample ($N = 2,471$) and an evaluation sample ($N = 1,178$). The proportion of each group that rendered effective service (base rate) was 0.73 for the development sample, 0.72 for the evaluation sample, and 0.72 for the total sample.

Statistical analysis in the development sample yielded an equation designed to generate probability of success estimates for all persons in the evaluation sample. The point-biserial cross-validity between predicted performance and actual performance was 0.315. Finally, for the sake of stability, the development and evaluation samples were combined and a multiple regression equation was developed on the total sample ($N = 3,649$). This equation was used to produce the probability of success estimates in the OFE-2 table.

It was recommended that the Navy Recruiting Command replace the OFE-1 table with the OFE-2 table. This was done and the OFE-2 table became operational on 1 October 1975. An ongoing NAVPERSRANDCEN research effort was designed to provide an updated version of the OFE table (OFE-3) based upon recent recruit input.

104. Sands, W. A. Screening male applicants for Navy enlistment (NPRDC TR 77-34).
San Diego, CA: Navy Personnel Research and Development Center, June 1977.
(AD-A040 534)

Recently, the Navy has experienced a premature attrition rate of more than one in every three newly enlisted personnel. The purpose of this effort was the development and evaluation of a new screening instrument that could be used by Navy recruiters in the field to estimate an applicant's probability of surviving the initial 2 years of service. Using this new instrument, the Prediction Of Enlisted Tenure - Two Years (POET-2) model, those applicants with a low probability could be screened out, resulting in a decrease in premature attrition. Essentially all nonprior service males enlisting during 1973 were included in the study ($N = 68,616$). Pre-data included: (1) aptitude test score, used to determine mental group, (2) years of school completed, (3) age at active duty base date, and (4) number of primary dependents. The dichotomous criterion was survival (72%) vs loss (28%) after a median 2 years of service. The model developed on the total sample evidenced a multiple point-biserial validity of .31. Double cross-validation evidence showed that the model will produce reasonably accurate and stable predictions. Management-oriented information was prepared that illustrated the various consequences of employing alternative cutting scores. This permitted examination of the tradeoffs involved in setting standards in the light of the current supply and demand picture for nonprior service enlisted males.

105. Schultz, D. G., & Siegel, A. I. Post-training performance criterion development and application: A selective review of methods for measuring individual differences in on-the-job performance. Wayne, PA: Applied Psychological Services, July 1961.

For several years, Applied Psychological Services has been carrying out research in the development and application of criteria for assessing the proficiency of Naval technicians in various technical specialties. Before undertaking additional work, it seemed wise to evaluate the current "state of the art" with respect to methods for the measurement of individual differences in on-the-job performance. This report considered recent progress in the area and attempted to point up a number of important issues which require investigation and clarification at this time.

Job performance appraisal techniques which have been used were discussed. These included production records, interviews and questionnaires, work sample and situation tests, appraisal of executive performance, and rating scales. Criterion analysis was reviewed in terms of intercorrelation and factor analysis, scaling, and reliability, including job performance changes over time.

Important current issue in the field of job performance measurement discussed were problems associated with the dimensionality of performance criteria, their selection and evaluation, their predictability, their ultimacy, and the influence of environmental factors.

It was concluded that there was need for an integrating conceptual framework to order and organize the field of measuring individual differences in on-the-job performance and to provide a more satisfactory basis for evaluating on-the-job performance.

106. Seymour, G. E., & Gunderson, E. Attitudes as predictors of adjustment in extremely isolated groups (NMNRU 70-37). San Diego, CA: Navy Medical Neuropsychiatric Research Unit, July 1970.

An analysis was conducted of the relative predictability, using attitude items as predictors, of five adjustment criteria in three occupational groups that participate in the U.S. Antarctic Research Program. Specificity of items for the various criteria and groups was characteristic. The Navy construction (Seabee) group was most predictable of the four specific criterion scores. This type of analysis helped to define the contributions of a particular set of attitude items to the prediction of specific aspects of adjustment for varied work roles in an unusual and extreme environment.

107. Sharp, L. H., Helme, W. H., & Boldt, R. F. Prediction of success in administration and machine accounting jobs (PRB Technical Research Note 94). Washington, D.C.: Personnel Research Branch, TAGO, DA, June 1958.

This study was one of a series evaluating aptitude area composites of the Army Classification Battery (ACB) for effectiveness in predicting performance in seven jobs in the Clerical Occupational Area. Scores on the ACB, on ACB test composites, and on final grades obtained in Army school courses prerequisite to assignment, were compared with supervisor and associate ratings of job performance of a total of 1301 men. Aptitude Area GT, General Technical, was as valid a selector for five of these jobs as was Aptitude Area CL, Clerical, currently in operational use, and slightly more valid for the two remaining jobs. A consideration of both job and prior school validity results indicated that substitution of GT as the selector for certain of these courses designated to train men in certain of these jobs was justified.

108. Sharp, L. H., Helme, W. H., & White, R. K. Prediction of success in selected electronics repair jobs (PRB Technical Research Note 92). Washington, D.C.: Personnel Research Branch, TAGO, DA, April 1958.

This study was one of a series evaluating the Army Classification Battery (ACB) for effectiveness in predicting performance in five electronics and electrical equipment repair jobs. Scores on the Army Classification Battery, on ACB test composites, and on final grades obtained in Army school courses prerequisite to assignment were compared with supervisor and associate ratings of job performance of a total of 747 men. Aptitude Area EL, Electronic, was the best available selector for four of the jobs, although, in general, validity was low. Aptitude Area MM, Motor Maintenance, was more valid than EL for the fifth job--that of Powerman--and a recommendation was made for an appropriate shift in Aptitude Area selector for the prerequisite course.

109. Siegel, A. I., & Wiesen, J. P. Experimental procedures for the classification of naval personnel (NPRDC TR 77-3). San Diego: Navy Personnel Research and Development Center, January 1977. (AD-A035 744)

Two concepts--miniature job learning and evaluation and assessment center methodology--were woven into a technique for evaluating and classifying personnel for technically oriented jobs. The concepts are presented and the resultant evaluative methodology described. Trial work indicated acceptable internal psychometric characteristics and considerable acceptability for the methods and approach.

110. Siegel, A. I., & Bergman, B. A. Nonverbal and culture fair performance prediction procedures. I. Background, test development, and initial results. Wayne, PA: Applied Psychological Services, Inc., June 1972.

The logic and initial results were described of a program in the development of unique measures for assessing the potential of "low aptitude" personnel for certain Navy rates. The logic was based on the conjecture that recruits who could learn a sample of the job requisites in a mini on-the-job training situation would demonstrate the same ability on the job. This hypothesis was held to apply regardless of any recruit's low score on the usual classification tests. The initial and criterion tests were described and the correlations among the mini job-learning test results and the usual Navy predictors were given. The results of a factor analysis of a questionnaire related to cultural deprivation were given, and the relationship of the derived cultural deprivation scores both to the usual Navy classification tests and the job learning tests were given.

111. Siegel, A. I., Bergman, B. A., & Lambert, J. Nonverbal and culture fair performance prediction procedures. II. Initial validation. Wayne, PA: Applied Psychological Services, Inc., September 1973.

The initial validation of a nonverbal, culture fair battery of tests for predicting performance of Navy machinist mates was described. The battery aspect of a job can serve as a predictor of ability to learn the job as a journeyman. The battery was administered to 50 black and 49 white recruits who were below the minimal acceptable score for admission to the machinist mate school training, as measured by the usual Navy written tests. These recruits were placed on the job and their level of competence was measured through work sample performance test methods nine months later. It was possible to acquire certain criterion data for 29 of the black and 25 of the white subjects. The results indicated that the performance battery correlated higher with the performance criterion than the usual Navy tests. In a considerable number of cases, the "low aptitude" sample performed better on the criterion tests than persons in a control sample who had surpassed the minimal acceptable Navy test scores and who had entered the specialty after attending the Navy school for machinist mates.

112. Siegel, A. I., & Leahy, W. R. Nonverbal and culture fair performance prediction procedures. III. Cross validation. Wayne, PA: Applied Psychological Services, Inc., March 1974.

A cross validation of findings relative to the value of a fair test concept was presented. The concept was based on the conjecture that persons who demonstrate the ability to learn a sample of the tasks of a job would, given appropriate on-the-job training, be able to achieve an absolute proficiency criterion of job success. An initial validation (conducted after the sample had 9 months job experience) had provided support for this contention. The cross validation (conducted after the sample had 18 months of job experience) similarly supported the contention. However, as anticipated, attenuation of predictive power was demonstrated in the 18-month cross validation follow-up. For the 9-month follow-up, the concept yielded discriminant functions that provided 74 percent correct classification. For the 18-month follow-up, 62 percent correct classification was demonstrated.

113. Siegel, A. I., Schultz, D. G., & Benson, S. Post-training performance criterion development and application: A further study into technical performance check list criteria which meet the Thurstone and Guttman scalability requirements. Wayne, PA: Applied Psychological Services, March 1960.

This study was one of a series by Applied Psychological Services in the development and application of criteria for post-training performance evaluation in the Navy. The specific purposes of this report paralleled those of an earlier report by Siegel and Benson. Their work utilized the skills involved in the work of aviation electronics technicians, whereas this study was based on the skills of the aviation machinist's mate and involved three phases. In the first phase, the hypothesis that skills are scalable in the same manner as attitudes and the sensory phenomena which have been previously scaled psychophysically was investigated. Three checklists were developed for the skills reflected by the tasks performed by the aviation machinist's mate. Two of these were shown to meet the criteria for a Thurstone equal-appearing interval scale, while the third did so only in a very rough sense. The two most discrepant of these lists were subjected to a Guttman analysis. One of the two checklists scaled according to Guttman's standards. The other, which had scaled only roughly in the Thurstone analysis, did not scale. Although these results suggested support for the hypothesis, the discrepant data raise some question as to the generality of the hypothesis as applied to aviation machinist's mates. Some possible explanations for the findings were discussed.

In the second phase, the hypothesis that the measured level of performance of the aviation machinist's mate would show a positive correlation with Naval attitudes as measured through an attitudinal inventory was investigated. The results suggested little, if any, support for this hypothesis. All obtained correlations between attitudinal inventory scores and the post-training performance evaluation scores were low. The third phase consisted first of an examination of all the intercorrelations among various predictors (attitudinal inventory scores, GCT, ARI, MECH, CLER, and final class average) and fleet performance; second, this phase involved the development of an equation to predict Scaled Technical Training Check List (STTCL) scores. Although no one variable had a high correlation with STTCL scores, the multiple correlation coefficient was found to be .42 for a combination of three attitudinal questionnaire subscores and a clerical aptitude test.

The results were felt to be consistent with the results of the Siegel-Benson study, with the possible exception of the lack of scalability for one of the checklists.

114. Siegel, A. I., & Benson, S. Post-training performance criterion development and application: Technical performance check list criteria which meet the Thurstone and Guttman scalability requirements. Wayne, PA: Applied Psychological Services, December 1959.

This report presents the results of five separate but related substudies: Substudy I and II investigated the hypothesis that skills were scalable in the same manner as are the attitudes and the sensory phenomena that had been previously scaled psychophysically. Three scales meeting the Thurstone criteria were developed for the skills underlying the tasks performed by the naval aviation electronics technician. It was also shown that these scales meet the Guttman criteria of scalability. Accordingly, within the framework of this study, this hypothesis can be regarded as substantiated.

Substudy III investigated the hypothesis that the measured level of performance of aviation electronics technicians will show a positive correlation with naval attitudes as measured through an attitudinal inventory. Little or no relationship was found to exist between naval attitudes and fleet proficiency, as measured in this study.

Substudy IV investigated the relationship between various "predictors" and the post-training performance effectiveness of naval aviation electronics technicians. Of the predictors investigated, no one predictor per se was found strong enough for practical individual prediction of fleet performance. A multiple R of .44 was achieved through a combination of General Classification Test scores and certain attitudinal variables.

Study V compared, in terms of maximum possible prediction, the power of the manifest structure analytic technique with the regression technique. The regression technique was found to be more powerful.

115. Sprunger, J. A., & Armore, S. J. Prediction of success in clerk jobs (PRB Technical Research Note 68). Washington, D.C.: Personnel Research Branch, TAGO, DA, December 1956.

An evaluation was made of the effectiveness of two-test composites of Army Classification Battery test scores for predicting job success of Clerks training in MOS 4405. The ACB test scores, the previous operational Aptitude Area scores, and scores on other potential composites were compared with ratings of on-the-job success of personnel in each of four job samples.

Two-test composites of the Army Clerical Speed Test plus the Arithmetic Reasoning Test (ACS + AR) and the Army Clerical Speed Test plus the Reading and Vocabulary Test (ACS + RV) were effective predictors of both Clerk success on the job and of Clerk course final grades.

116. Sprunger, J. A., & Armore, S. J. Prediction of success in cook jobs (PRB Technical Research Note 67). Washington, D.C.: Personnel Research Branch, TAGO, DA, December 1956.

An evaluation was made of the effectiveness of two-test composites of Army Classification test scores for predicting job success of cooks trained in MOS 1824. The ACB test scores, and previous operational Aptitude Area III scores, and scores on other potential composites were compared with ratings of on-the-job success of personnel in two samples of 237 each. Although there are several potentially good predictors of course success, there is a need for a better predictor of on-the-job success than was identified in this study.

117. Standlee, L. S., & Abrahams, N. M. Selection of Marine Corps Drill Instructors (NPRDC TR 80-17). San Diego, CA: Navy Personnel Research and Development Center, March 1980. (AD-A082 966)

The purpose of this effort was to assist the Marine Corps in more accurately predicting the success of prospective drill instructors. Students entering Drill Instructor (DI) school ($N = 759$) were administered an experimental test battery that covered both intellectual and motivational factors. Analyses of responses showed that a composite score of volunteer status, General Classification Test score, and level of education, and a Biographical Questionnaire score were predictive of performance in DI school. Performance in DI school was the best single predictor of performance on the job.

118. Swanson, L., & Anderson, A. V. Peer ratings as an immediate criterion of Sonarman performance. II. Relationship between peer ratings and shipboard rating measures (PRFASD Report No. 85). San Diego, CA: Naval Personnel Research Field Activity, August 1955.

The purpose of this study was to evaluate the use of peer ratings obtained at the Fleet Sonar School, San Diego, as an immediate criterion of sonarman performance by relating them to shipboard performance as measured by the Shipboard Rating Scale for Sonarmen. Reliabilities of peer ratings were determined by the split-half method and corrected by the Spearman-Brown formula. Ratings by peers, rankings by instructors, and school measures were related by correlational methods to sonarman shipboard performance as measured by supervisors' ratings obtained with the Shipboard Rating Scale for Sonarmen. Selection measures were also included in the analysis.

At the time of shipboard evaluations, all of the men in the sample had had 6 or more months of shipboard experience following graduation from the Basic Sonarman Course 560 at the Fleet Sonar School, San Diego. Shipboard rating scales were obtained from February 1953 to February 1955 on a sample of 203 subjects who had no fleet experience before entering sonar school.

For peer ratings, obtained on 82 groups of from 7 to 13 students at the end of sonar school training, the median reliability was .87. Peer ratings appeared to be related to school performance measures to a moderate degree. The correlation between peer ratings and shipboard performance as measured by the Shipboard Rating Scale for Sonarmen total score was low, but significant at the 1 percent level; the probable magnitude of the true relationship was in doubt because of the unknown reliability of the shipboard criterion.

Because the demonstrated relationships between peer ratings and performance in either school or in the fleet are so low as to have little practical value, peer ratings were not recommended for operational use. It was recommended that no further evaluation of peer ratings as an immediate criterion of sonar performance should be attempted until more adequate measures of shipboard Sonarman performance were available.

119. Swanson, L. Relationships among aptitude, school and shipboard measures for Sonarmen: An analysis with revised criterion measures. San Diego, CA: Naval Personnel Research Field Activity, December 1955.

The primary purpose of this study was to validate current selection requirements for sonar school against school grades and shipboard performance as measured by a shipboard rating scale. A secondary purpose was to investigate the relationships between a group of tests, experimentally administered at the beginning of sonar school training, and school performance. For two Key West groups and one San Diego group, correlations among selection tests, school grades, and shipboard performance were determined. For a second San Diego group, correlations among selection tests, experimental tests, and school performance were computed. Biserial and multiple correlations between predictor tests and the graduate-drop criterion were determined. Analysis of cutting scores on the General Classification Test (GCT), Arithmetic Test (ARI), and the Electronics Technician Selection Test (ETST), were made to determine their value for selection of sonar students.

Findings were that (1) GCT and ARI are generally significantly related to phase and final grades in sonar school; (2) the Sonar Pitch Memory Test is a good predictor of Sound Recognition Group Trainer (SRGT) grade in sonar school; (3) in this restricted population none of the selection tests is significantly related to shipboard performance; (4) phase and final school

grades are significantly related to performance aboard ship in two of the three groups where this type of information is available; and (5) of the experimental variables, ETST is the best predictor of the graduate-drop criterion in sonar school.

120. Thomas, P. J. An evaluation of methods for predicting job performance of Personnelmen (Technical Bulletin STB 72-4). San Diego, CA: Naval Personnel and Training Research Laboratory, September 1971.

Attention has been focused upon Navy ratings that represent contact points between Navy policies and the enlisted men and their dependents. The Personnelman (PN) rating was the subject of one recent study in which PN selection test scores were found to correlate satisfactorily with school grades. The purpose of this follow-up study was to determine: (1) correlations between selection test scores and on-job performance measures; and, (2) if various experimental tests administered to PN students in school are related to performance in the PN rating. Job performance evaluations were obtained for samples of PNs six months after graduation from A School from the Report of Enlisted Performance Evaluation (NAVPER 792) and from an experimental Personnelman Supervisor's Questionnaire. Basic Test Battery (BTB) scores, experimental test data, and school grades were validated against the criteria of job performance. Comparisons were made among the four samples of school graduates and between men who entered the schools from the fleet and directly from recruit training.

Measures obtained at the end of school training, i.e., Peer Ratings, Instructor's Ratings, and Final School Grades (FSB), were substantially related to job performance (r_s ranged from .23 to .35 with the total sample and achieved .60 in one school). The BTB tests used for assignment to PN school, GCT and ARI, also were significantly correlated with job performance with r_s ranging from .14 to .19. The CLER Test was virtually unrelated to PN job performance. None of the experimental memory tests and neither of the vocational interest scales were consistently correlated with the criteria to a significant degree.

121. Vineberg, R., & Joyner, J. N. Performance of men in different mental categories: 2. Assessment of performance in selected Navy jobs (Tech Rpt 78-1). Carmel, CA: Human Resources Research Organization, September 1978.

Worker-oriented and job-oriented supervisor rating instruments that could be used to evaluate the elements of behavior and performance of tasks

in a job were developed. The job performance of persons in Mental Categories 1-4 was assessed in a variety of Navy jobs in pay grades E3-E5. There is no clear evidence that persons in lower mental categories were less effective either in the rated quality of their performance or in the number and characteristics of the duties they performed. Supervisors perceived the most effective job incumbents in pay grades E3 and E4 to be persons in either the highest or lowest mental categories and the most effective incumbents in Grade E5 to be persons in the lower mental categories. This pattern may be interpreted in terms of (1) the relative importance of technical factors and non-technical factors in job performance and their influence on ratings of performance, and (2) selective processes which favor the acquisition and retention of effective performers in the lower mental categories.

122. Vineberg, R., & Taylor, E. N. Performance in four Army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance (HumRRO TR-72-22). Alexandria, VA: Human Resources Research Organization, August 1972.

To provide information on performance and characteristics of effective and ineffective marginal personnel in the Army, a study has been made of approximately 1500 men with experience ranging up to 20 years in four different Army MOSs. The study included a group of men with Armed Forces Qualification Test scores in the marginal range and a comparison group of men in the same jobs, but in the upper AFQT levels. This report, the third in a series, described the bulk of the major study findings, including comparisons of the performance of men in different mental categories with different amounts of job experience, comparisons of the performance of special subgroups (Negroes and Caucasians, inductees and enlistees, and men with formal and on-the-job training), an analysis and definition of acceptable performance, and a procedure for using job knowledge tests to screen ineffective performers.

123. Vineberg, R., & Taylor, E. N. Performance in four Army jobs by men at different aptitude levels: 4. Relationships between performance criteria (HumRRO TR 72-23). Alexandria, VA: Human Resources Research Organization, August 1972.

A study was made of approximately 1800 men with experience ranging to 20 years in five different Army MOSs to provide information about the performance and characteristics of effective and ineffective marginal personnel in the Army. The study included a group of men with Armed Forces Qualification Test scores (AFQT) in the marginal range and a comparison group of men in the same jobs, but in the upper range of AFQT scores. Performance was measured by intensive job sample tests, job knowledge tests, and supervisor ratings. Biographical questionnaires, a battery of published and experimental tests, and Army records provided information about background, personal characteristics, and military experiences. This report, the fourth in a series

presenting the extensive data and analyses, examined the determinants of job behavior and described the relationships among the three performance criteria used in the study: job sample tests, job knowledge tests, and supervisor ratings.

124. Weeks, J. L., Mullins, C. J., & Vitola, B. M. Airman classification batteries from 1948 to 1975: A review and evaluation (AFHRL-TR-75-78). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1975.

From 1948 to 1975, the United States Air Force employed ten different multiple batteries for the purpose of either classifying or selecting and classifying nonprior service enlistees. Each of the different batteries was described and evaluated in terms of standardization, reliability, and validity.

125. Wilcove, G. L., Thomas, P. J., & Blankenship, C. The use of preenlistment variables to predict the attrition of Navy female enlistees (NPRDC SR 79-25). San Diego, CA: Navy Personnel Research and Development Center, September 1979.

Although the attrition rate for first-term enlisted women has been decreasing, it is still unacceptable to the Navy. The purpose of the present study was to conduct the exploratory research necessary to develop a questionnaire for screening female applicants. Attrition factors were identified from interviews and from research on turnover, mental health, sex roles, and vocational choice. These factors were used to construct two experimental questionnaires (QUEST 1 and QUEST 2). One or the other of the questionnaires was administered to each of 997 female recruits. Empirical keying was employed to create "response-option" scales to predict attrition, which were then validated. Thirty-eight items were found to be significantly related to attrition, and an estimated cross-validation R of .30 was obtained in a multiple-regression analysis. A response-option scale constructed from the unique QUEST 1 items yielded the highest validation correlation, - .25. The 18 items should be evaluated further, and psychometrically and legally inappropriate items should be dropped. An attempt should then be made to determine whether remaining items improve prediction over and above the Armed Services Vocational Aptitude Battery, and a similar determination should be made for a "second generation" response-option scale.

126. Wiley, L. N. Across-time prediction of the performance of airman administrators and mechanics (AFHRL-TR-74-53). Lackland Air Force Base, TX: Occupational Research Division, Air Force Human Resources Laboratory, July 1974.

It was found that supervisors' ratings of Administration Specialists' and Aircraft Mechanics' job performance were predictable across time.

Airmen in duty AFSCs 702X0/70490 and 431X1/43190 were rated on overall job performance and 65 traits. After 2 years for mechanics and 3 years for administrators, the available airmen were located and rerated. More than half were rated by two supervisors on each occasion, which permitted testing the agreement between raters for airmen at the same skill levels. At least 16 percent of the Time 2 performance variance was predictable from trait ratings with multiple Rs from .40 to .47. The first overall performance ratings made less prediction than did the 65 trait ratings taken as a whole. The results helped to support earlier findings, on samples which included these airmen, that the traits important for the performance of mechanics differed somewhat from the traits important for administrators, and that skill levels within ladders differed in their trait requirements. The traits used were statements of consistent work behaviors, as distinguished from vague generalizations.

127. Wiley, L. N. Airman job performance estimated from task performance ratings (AFHRL-TR-76-64). Lackland Air Force Base, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, October 1976.

An experiment was conducted to determine whether a job performance criterion could be developed from averaging airman performance of separate tasks. Airmen who had completed job inventories in the supply field, AFSCs 645X0 and 647X0 from all commands and locations in 1967-1968 were rated by two supervisors in a confidential study. The immediate supervisor and another supervisor were demanded, with complete rating data and an acceptable job inventory. Despite stringent stipulations, 244 airmen, representing all supply levels and locations, were rated by two supervisors, providing 488 independent sets of ratings. These included an overall rating, ratings on 65 work behavioral traits, performance ratings on all tasks the supervisor was certain the airman performed, and a time-to-train rating on each task in the inventory. The mean task performance rating and the mean task trainability rating were computed. The three criteria of overall performance rating, mean task performance rating, and mean task trainability rating were compared through cross-correlations and cross-regressions, using both the 244 airmen data and the maximum set of 488 observations. The cross-validity of the overall criterion was .58, compared with .56 for the mean task performance rating and .43 for the mean task trainability rating. The regressions showed large contributions from the work behavior ratings, but from the data of record, including grade and job difficulty indices, the contributions were nonsignificant. The mean task performance rating was not cost effective for lower level airmen from the standpoint of rating time consumed. However, the possibility remained open that it might be cost effective for upper level airmen when combined with securing information about the requirements of unusual tasks.

128. Wiley, L. N., & Cagwin, L. P. Comparing prediction of job performance ratings from trait ratings for aircraft mechanics and administrative airmen (AFHRL-TR-68-108). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory, October 1968.

Supervisors in all commands rated aircraft mechanics on overall job performance and on 65 work-related traits. Of 1,290 ratees, there were 852 who were rated by each of two supervisors, providing samples of 83 in DAFSC 43131, 418 in DAFSC 43151, 274 in DAFSC 43171, and 77 in DAFSC 43190. Trait predictions of overall performance yielded R's ranging from .78 to .94, and cross-validation R²'s from .33 to .86. Interpretations involved comparisons with previous findings obtained from ratings on administrative airmen. The analyses added confirmation in a different career ladder of most of the administrative ladder findings and suggested that there are some areas where the interpretations cannot be generalized from one work situation to another. It was concluded that any supervisor should be able to make this type of rating if given opportunity to observe the man. Particular attention should be given to the opportunity of supervisors to observe men.

129. Wiley, L. N. Describing airman performance in the administrative career ladder by identifying patterns of trait ratings (PRL-TR-66-13). Lackland Air Force Base, TX: Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, November 1966.

Trait ratings were used to account for the variance in airman performance reports and in overall experimental performance ratings. Airmen in the administrative career ladder, DAFSCs 70230, 50, 70, and 70490, across all commands, were rated by supervisors on overall performance and on 65 traits. Current overall airman performance reports (APRs) were obtained from base records. Among the 2,606 sets of ratings with complete data, 1,083 individuals were evaluated twice, representing personnel rated by two supervisors. Broken down by skill levels, the smallest N was 140, for 9-level men who had been rated twice. Using data undifferentiated by skill, in which a man might appear twice if so rated, trait ratings accounted for 70 percent of the variance in experimental performance ratings and about 43 percent of the variance in APRs, after grade was removed as a predictor. When data were sorted by skill level, prediction held up in all skills except DAFSC 70270, where it dropped to 60 percent. Patterns of traits which were more predictive of performance in one skill level than another were found, and these patterns could be sensibly interpreted in terms of the expected demands of the jobs. In a cross-validation against different raters, the predictive advantage of selected patterns was found to be statistically significant for the 5-, 7-, and 9-skill levels. The study was discussed in terms of its implications for criterion development, particularly in respect to its place in the sequence of current criterion research studies.

130. Wiley, L. N. Familiarity with subordinates' jobs: Immediate versus secondary supervisors (AFHRL-TR-75-7). Lackland Air Force Base, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, June 1975.

A test was made of the hypothesis that only immediate supervisors know enough about their subordinates' job activities to render job performance ratings. Pairs of supervisors who rated the quality of performance of Supply airmen had identified themselves as immediate supervisors and other-than-immediate supervisors. These pairs, working independently, rated the same airmen on how well they performed individual tasks. Each supervisor was asked to rate each task that he was sure the subordinate did, but he was not told which tasks the subordinate had identified. The selection of tasks were tallied against the responses made by the incumbents on the same inventory. An incumbent's responses were relative time spent ratings. Tasks were classified by a scale of percent time spent, and two supervisory levels were compared in terms of percentage of tallies ("agreements") with the incumbents. The tallies were greater for tasks on which the airmen spent more time, but there was no detectable difference between immediate and other supervisors. It was concluded that in the Inventory Management, DAFSC 645X0, and Materiel Facilities, DAFSC 647X0, career ladders, at least, it was possible to obtain other supervisors who were as familiar with their subordinates' jobs as "immediate" supervisors.

131. Wiley, L. N. Ratings of first-term airmen on supervisory potential and technical competence in AFSCs 462X0 and 812X0 (AFHRL-TR-75-56). Lackland Air Force Base, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, October 1975.

A study was undertaken to determine whether supervisors could rate the potential of first-term airmen to become supervisors. Ratees were 313 Weapons Mechanics, AFSCs 462X0 and 461X0, and 421 Law Enforcement Specialists, AFSCs 812X0 and 811X0, who were rated on 3 criteria and 30 job behavioral traits by their supervisors in CONUS. The criteria of (1) supervisory potential, (2) technical competence, and (3) desirability as a reenlistee were predicted from their correlation with the 30 trait ratings by linear regression techniques. The aim was to see if the Weapons Mechanic and Law Enforcement specialties differed in their supervisory trait requirements, and if supervisory potential is distinguishable from technical competence. The first two criteria correlated .89 with each other, while the criterion of desirability as a reenlistee had to be discarded because it was not uniformly interpreted by the raters. Both technical competence and supervisory potential were highly predictable from trait ratings, 86 and 84 percent, respectively. However, through direct examination of the data and the supervisors' comments, it was concluded that the supervisory requirements of the two specialties actually differ, and that technical competence was an element of supervisory potential, a necessary but not sufficient attribute of a future supervisor.

132. Wiley, L. N. Relation of job qualification ratings to performance ratings of basic training instructors (PRL-TDR-64-21). Lackland Air Force Base, TX: Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, July 1964.

Among the many studies of selection and classification instruments, few have shown high relationship between selection tests and job performance ratings. It was hypothesized that some of the prediction failures could arise from mixing jobs with dissimilar requirements in the criterion data. The job of Tactical Instructor (TI) was selected to test whether a job requiring all incumbents to perform the same tasks would yield reliable performance data that would be predictable from a battery of qualifications ratings. Fifty-five NCO supervisors rated 527 TIs on overall job performance and on 45 job qualification characteristics. By multiple regression techniques, it was found that characteristics ratings accounted for 75 percent of the variance in the overall ratings. Three months later 53 of the supervisors rerated 482 TIs. The correlation between the two ratings (reliability) was .72. Overall ratings of 309 TIs by 12 supervisory lieutenants correlated .63 with the reratings. Ratings of the 45 characteristics accounted for 60 percent of the rerate variance and 50 percent of the variance in lieutenants' ratings. The findings were consistent with the hypothesis that some of the unpredictability of job performance ratings may be due to mixing dissimilar jobs in collecting criterion data.

133. Wiley, L. N., & Hahn, C. P. Task level job performance criteria development (AFHRL-TR-77-75). Brooks AFB, TX: Air Force Human Resource Laboratory, and Washington, D.C.: American Institutes for Research, December 1977.

This study investigated the possibilities for improving the identification of the requirements for jobs by studying performance of job incumbents on separate tasks. Three specialties were selected for study: 291X0, Telecommunications Operations Specialist; 304X4, Ground Radio Communications Equipment Repairman; 431X1C, Aircraft Maintenance Specialist, single- and dual-engine jet. Incumbents, peers, and supervisors rated the performance of the incumbents on a selected set of tasks. In addition, job inventories and an experimental test battery were administered to the incumbents. The battery included 11 short experimental cognitive tests, a Biographical Inventory, the Vocational Interest-Career Examination (VOICE), and a 43-item Job Satisfaction Information blank. Data of record were also obtained from Air Force files to provide such items as incumbent grade, service time, sex, education at enlistment, and Aptitude Index scores. Correlations were run between raters, correlating performance on separate tasks, and between raters, correlating performance on 6 overall dimensions of appraisal. Cross-rater reliabilities were low, but significant, on task assessments, and in the $r = .40$ range on overall

ratings. Similarly low correlations were found for nontask predictors, such as grade, service time, and Aptitude Indexes. All types of obtained measures, except data on the origins of training and on task performance satisfaction, were put into regression problems to account for the 6 overall performance ratings made by peers and supervisors. The data suggest that different factors were important for different kinds of work, and for different dimensions of performance appraisal. Of all the many findings of the study, by far the most enlightening was that difficult tasks (in terms of learning time) were better measured on performance. This arose from less use of the top of the rating scale, and it produced lower performance appraisals from the group (AFSC 304X4) selected by the Air Force for having the highest aptitude scores. Should subsequent analyses prove that this finding also applies to job ratings within AFSCs, the result would have implications for Air Force job performance appraisal.

134. Willemin, L. P., and Karcher, E. K., Jr. Development of combat aptitude areas (PRB Technical Research Report 1110). Washington, D.C.: Personnel Research Branch, TAGO, DA, January 1958.

Existing Army Classification Battery tests and Aptitude Area composites have since 1949 been shown to be consistently valid for assigning enlisted men into a multitude of technical, common specialty, and support jobs but less valid for assigning to combat. In studies conducted in the Arctic in 1949 and 1950, in Korea in 1951 and 1953, and in a training-maneuvers situation in 1955 and 1956, promising new tests measuring vital personality and interest aspects of successful combat potential were developed to predict combat, maneuver-garrison and AIT criteria. These were then refined and given the necessary experimental tryout in establishing their utility as Army classification procedures. As a result of this research, two new tests were introduced into the Army Classification Battery. They formed part of two new Aptitude Areas for classifying to the combat arms. Aptitude Area IN, consisting of the Classification Inventory and the Arithmetic Reasoning Test, was installed as the best available test composite for classifying to Infantry. Aptitude Area AE, consisting of the General Information Test and the Automotive Information Test, was found best for Artillery, Armor, and Combat Engineer assignment.

135. Willemin, L. P., de Jung, J. E., & Katz, A. Prediction of enlisted personnel under conditions of extreme cold (PRB Technical Research Report 1113). Washington, D.C.: Personnel Research Branch, TAGO, DA, September 1958.

Although the magnitude of validity revealed for the predictors of the cold weather criterion was not as great as previously obtained against combat and garrison-maneuver criteria, the pattern of validity among test variables was a familiar one. The new combat aptitude area composites, the two new ACB tests contained therein (CI and GIT), the Arctic BIB, and the Shop Mechanics Test of the ACB were the best predictors of peer rankings of tentmates with respect to COLD BAY maneuver performance. Background variables of age and grade also showed significant variation with this criterion. The research turned up no new leads for combat prediction.

136. Willemin, L. P., Birnbaum, A. H., Rosenberg, N., & White, R. K. Validation of potential combat predictors in overseas maneuvers (PRB Technical Research Note 80). Washington, D.C.: Personnel Research Branch, TAGO, DA, August 1957.

This study was one of a series to improve effectiveness of the Aptitude Area system of personnel classification and assignment for the combat arms. In earlier studies, a large number of test materials had been prepared for later use in a large scale study following recruits through training and through performance in maneuvers overseas.

The purpose of the study was to furnish Infantry, Artillery, Armor, and Combat Engineer research information for use in selecting new Combat Arms Aptitude Areas based upon the validity of ACB tests and a group of experimental predictors. Testing was accomplished in the 10th Infantry Division at the start of the training cycle for 1642 enlisted men later trained in Combat Arms Military Occupational Specialties. Criterion ratings of estimated combat potential were collected after overseas maneuvers, one year after testing.

The most valid composites included both ACB tests and experimental predictors. The pattern of validity results for Combat Engineer differed appreciably from those for the other Combat Arms. Nevertheless, the test composites chosen as most valid in each of the Combat Arms were more valid than the present Combat Aptitude Areas, not only for the branch in which they were chosen, but in general for the other branches also, including Combat Engineer.

137. Willemin, L. P., Birnbaum, A. H., & Rosenberg, N. Validation of potential combat predictors: Research plan of longitudinal study (PRB Technical Research Note 73). Washington, D.C.: Personnel Research Branch, TAGO, DA, June 1957.

This study was one of a series to improve effectiveness of the Aptitude Area system of personnel classification and assignment for the combat arms. In early studies a large number of test materials had been prepared for later use in a large scale study following recruits through training and through performance in maneuvers overseas. This study described the use of over 4000 members of a Gyroscope unit--the 10th Infantry Division--in designing a suitable research plan, administering 17 experimental instruments, and collecting information on training and on-the-job performance preliminary to identification of new Combat Aptitude Areas for selection to the combat arms. Results of research findings will be reported separately for the combat branches involved--Infantry, Artillery, Armor, and Engineer branches.

138. Wilson, C. L., Mackie, R. R., & Buckner, D. N. Research on the development of shipboard performance measures: Part IV. A comparison between rated and tested ability to do certain job tasks. Los Angeles, CA: Management and Marketing Research Corporation, February 1954.

Research into the problem of evaluating the shipboard performance of Navy enlisted personnel included the development of four types of performance measures: Performance Rating Scales, Performance Check Lists, Job Sample Performance Tests, and Written Job Knowledge Tests. The descriptions of these measures and the results of correlational analyses of total scores made on them by Electrician's Mates (EMs) and Enginemen (ENs) serving aboard submarines had been included in Parts I, II and III of the final report.

To supplement the total score analyses, additional correlational studies were made of scores on similar items of three of the performance measures. Below is a brief description of the item-score correlational analyses performed:

1. Check List Task-Items--Job Sample Tests. Scores on selected task-items of the check lists were correlated with scores on individual job sample tests. The task-items selected were those that appeared to reflect the same knowledge or skills as the job sample test task and could be expected, therefore, to yield reasonably valid ratings of men's abilities to perform the job sample test task.
2. Check List Task-Items--Written Job Knowledge Test Items. Scores derived from check list ratings of men's abilities to perform a particular task were correlated with scores these same men made on written test items about the task.

3. Written Test Items - Job Sample Tests. Scores made on a job sample test constructed around a particular task were correlated with scores the same men made on written test questions about that task.

The correlations obtained from these item score analyses ranged from zero to moderately high values. While many were significantly greater than zero, on the average they were quite low indicating that ratings of men's abilities to perform a particular task were not in substantial agreement with the same men's scores on tests of their ability to perform the task. The correlational analyses of total check list scores, reported in Part III of this final report, led to a similar conclusion. The results of this study also revealed substantial differences between the ability actually to perform a particular task and the ability to answer written questions about the same task. These results support other observations that scores on written tests cannot always be accepted as valid indications of men's abilities to perform practical tasks.

139. Wilson, C. L., Mackie, R. R., & Buckner, D. N. Research on the development of shipboard performance measures: Part III. The use of performance check lists in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, CA: Management and Marketing Research Corporation, February 1954.

Part II of the final report on research conducted to investigate methods of measuring the shipboard performance of Navy enlisted men describes the development and experimental use of performance check lists in evaluating the performance of Electrician's Mates and Enginemen serving aboard submarines.

In addition to information concerning the development and administration of the check lists, and estimates of inter-rater agreement in using them, this report contains discussions of the relationships between scores on the check lists and scores on the performance rating scale and practical performance tests, which are described more fully in Parts I and II of this Final Report.

140. Wilson, C. L., Mackie, R. R., & Buckner, D. N. Research on the development of shipboard performance measures: Part II. The use of a performance rating scale in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, CA: Management and Marketing Research Corporation, February 1954.

A performance rating scale that included 10 traits reflecting various technical and non-technical aspects of Navy shipboard performance was developed and used to evaluate the performance of Electrician's Mates (EMs) and Enginemen (ENs) serving aboard submarines in the Atlantic and Pacific Fleets. Analysis of results indicated that officers and petty officers using the scale tended:

(1) to agree with one another when they evaluated the same men; (2) to be consistent in their own evaluations from one time to the next; (3) to discriminate reliably among men of the same pay grade; (4) to differentiate, to an appreciable degree, the technical from the adjustive aspects of shipboard performance.

In addition, a factorial analysis indicated that: At least two board "factors" of shipboard performance--one representing technical skill, and the other, adjustment to Navy life--accounted for most of the inter-correlations among traits; the traits representing the technical side of performance correlated moderately high with independent measures of technical skill, but the traits representing the adjustment side of performance were not related appreciably to any other measures obtained.

As a part of this overall research project practical performance tests and performance check lists for EMs and ENs were also developed. Relationships between these two measures and the rating scale were reported and discussed in this report.

141. Wilson, C. L., & Mackie, R. R. Research on the development of shipboard performance measures: Part I. The use of practical performance tests in the measurement of shipboard performance of enlisted naval personnel. Los Angeles, CA: Management and Marketing Research Corporation, November 1952.

Research was conducted to determine whether objective and reliable measures could be developed to evaluate shipboard performance of Navy enlisted men. Adequate measures of performance are necessary to determine proper qualifications for advancement and to determine the effectiveness of selection and training programs. A series of Practical Performance Tests, designed to measure the practical factors of shipboard performance, were developed. These tests were administered to Electrician's Mates and Engine-men serving aboard submarines. They were shown to be valuable additions to existing performance measures.

Information about the usefulness and importance of performance tests was presented in this report. The reliabilities and interrelationships of the tests were discussed, and observations were made on the construction of performance tests. Results of correlational studies with other measures of shipboard performance were also given. Parts II through V of this final report will describe the development of shipboard Performance Check Lists and a shipboard Performance Rating Scale, and the administration of experimental aptitude tests to candidates at the Enlisted Submarine School in New London.

142. Wollack, L., & Kipnis, D. Development of a device for selecting recruiters.
Washington, D.C.: U. S. Naval Personnel Research Field Activity, March
1960.

The purpose of this study was to develop an objective instrument or test battery that would significantly increase the probability of selecting successful Navy recruiters. Experimental tests consisting of the Kuder Preference Record, the Navy Knowledge Test, the Career Motivation Survey, the Career Preference Scale, and the Sports Inventory were administered to 410 men considered by their supervisors to be effective or ineffective recruiters. Items and tests that discriminated effective from ineffective recruiters were cross-validated upon a second sample of 260 recruiters. In addition to the experimental battery given to the validation sample, four measures of verbal fluency were administered to the follow-up group. Testing was carried out at Personnel Man Class "C" Schools and judgments of recruiting effectiveness were obtained from commanding officers after the men in the cross-validation sample had been on the job for approximately one year.

It was found that the Persuasive Scale and the Scientific Scale of the Kuder predicted evaluations of recruiting effectiveness (r 's of .24 and -.17, respectively). The item analysis key for the Career Motivation Survey also predicted this criterion ($r = .13$). None of the remaining tests correlated significantly with evaluations of recruiting effectiveness. The Kuder Persuasive Scale appeared to have value as a screening instrument in the selection of recruiters. Its contribution in operational use would depend upon the available numbers of applicants who met current eligibility requirements. Given a sufficient applicant pool, elimination of men with low Persuasive Scale scores could be expected to elevate the quality of input of men into recruiting assignments.

143. Zeidner, J., Harper, B. P., & Karcher, E. K. Reconstitution of the Aptitude Areas (PRB Technical Research Report 1095). Washington, D.C.: Personnel Research Branch, TAGO, DA, November 1956.

The ten Aptitude Areas introduced in 1949 combined Army Classification Battery test scores on the basis of information then available. Sufficient additional information on the effectiveness of composites of ACB tests was accumulated to require review of the classification system.

Information from forty-two research studies on the effectiveness of composites for predicting success in Army school training and in Army jobs was used in constituting new Aptitude Areas. The seven new Aptitude Areas had the operational and technical advantages of combining only two ACB tests at a time, using each ACB test a less number of times, identifying the highest aptitudes of a greater percentage of men, and differentiating better the levels of aptitude in each man.

DISTRIBUTION LIST

Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics)
Assistant Secretary of the Navy (Manpower and Reserve Affairs)
Director of Manpower Analysis (ODASN(M))
Deputy Assistant Secretary of the Navy (CPP/EEO)
Chief of Naval Operations (OP-01), (OP-11), (OP-12) (2), (OP-13), (OP-14), (OP-15), (OP-115) (2), (OP-135), (OP-140F2), (OP987H)
Chief of Naval Material (NMAT 0722), (NMAT 08), (NMAT 08L)
Chief of Naval Research (Code 200), (Code 440) (3), (Code 442), (Code 448)
Chief of Information (OI-213)
Chief of Naval Education and Training (02), (N-5)
Chief of Naval Technical Training (016)
Commandant of the Marine Corps (MPI-20)
Commander in Chief, United States Naval Forces, Europe (2)
Commander Fleet Training Group, Pearl Harbor
Commander Naval Military Personnel Command (NMPC-013C), (NMPC-5), (NMPC-48)
Commander Navy Recruiting Command
Commanding Officer, Naval Aerospace Medical Institute (Library Code 12) (2)
Commanding Officer, Naval Education and Training Support Center, Pacific
Commanding Officer, Naval Regional Medical Center, Portsmouth, VA (ATTN: Medical Library)
Commanding Officer, Naval Training Equipment Center (Technical Library)
Commanding Officer, Office of Naval Research, Western Regional Office
Director, Naval Civilian Personnel Command
Director, Training Analysis and Evaluation Group (TAEG)
Officer in Charge, BUMED East Coast Equal Opportunity Program Detachment
Officer in Charge, BUMED West Coast Equal Opportunity Program Detachment
Officer in Charge, Naval Occupational Development and Analysis Center
President, Center for Naval Analyses
Superintendent, Naval Postgraduate School
Superintendent, U.S. Naval Academy
Commander, U.S. Army Soldier Support Center, Fort Benjamin Harrison (Human Dimensions Division)
Commander, Army Research Institute for the Behavioral and Social Sciences, Alexandria (PERI-ASL)
Headquarters Commandant, Military Enlistment Processing Command, Fort Sheridan
Director, U.S. Army TRADOC Systems Analysis Activity, White Sands Missile Range (Library)
Chief, Army Research Institute Field Unit--USAREUR (Library)
Chief, Army Research Institute Field Unit, Fort Harrison
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Manpower and Personnel Division)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Scientific and Technical Information Office)
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base (Technical Training Branch)
Commander, Air Force Human Resources Laboratory, Williams Air Force Base (AFHRL/OT)
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base (AFHRL/LR)
Commander, 314 Combat Support Group, Little Rock Air Force Base (Career Progression Section)
Commander, Federal Computer Performance Evaluation and Simulation Center

Director, Defense Equal Opportunity Management Institute, Patrick Air Force Base
Commandant Coast Guard Headquarters
Commanding Officer, U.S. Coast Guard Institute
Superintendent, U.S. Coast Guard Academy
President, National Defense University (3)
Director, Office of Personnel Management
Defense Technical Information Center (DDA) (12)